1984

Data dictionary systems and their role in information resources management.

Robertson, Debra Lynne

Monterey, California. Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

DATA DICTIONARY SYSTEMS AND THEIR
ROLE IN INFORMATION RESOURCE MANAGEMENT

by

Debra Lynne Robertson

March 1984

Thesis Advisor: Daniel R. Dolk

Approved for public release, distribution unlimited

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Data Dictionary Systems and Their Role in Information Resource Management		5. TYPE OF REPORT & PERIOD COVERED Master's Thesis March 1984
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Debra Lynne Robertson		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Postgraduate School Monterey, California 93943		12. REPORT DATE March 1984
		13. NUMBER OF PAGES 92
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

data dictionary systems (DDS), data dictionary/directory systems (DD/DS), information resource management (IRM)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The explosive proliferation of computers has led to the increasing importance of developing and implementing various management concepts for effective and efficient operation and control. The complex data processing environment of today cannot be handled by hardware alone, but require an information system composed of hardware, software, data, personnel

(20. ABSTRACT Continued)

and procedures. The vast storage capabilities of modern equipment has led to the development of databases for more effective and efficient use of memory capacity. The increasing importance of software and the cost of developing and maintaining it demands more and better management, giving rise to the software life cycle concept. With the automation of the functions of an organization, data and information become critical organizational resources. Information Resource Management provides effective and efficient management and control of these information resources. A key component in this management and control is the Data Dictionary System.

Data Dictionary Systems and Their Role in Information Resource Management

by

Debra Lynne Robertson
Lieutenant, United States Navy
B.S., University of California, 1977

Submitted in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN INFORMATION SYSTEMS

from the

NAVAL POSTGRADUATE SCHOOL
March 1984

ABSTRACT

The explosive proliferation of computers has led to the increasing importance of developing and implementing various management concepts for effective and efficient operation and control. The complex data processing environment of today cannot be handled by hardware alone, but require an information system composed of hardware, software, data, personnel and procedures. The vast storage capabilities of modern equipment has led to the development of databases for more effective and efficient use of memory capacity. The increasing importance of software and the cost of developing and maintaining it demands more and better management, giving rise to the software life cycle concept. With the automation of the functions of an organization, data and information become critical organizational resources. Information Resource Management provides effective and efficient management and control of these information resources. A key component in this management and control is the Data Dictionary System.

TABLE OF CONTENTS

LIST OF FIGURES

I. INTRODUCTION

The "Information Revolution" of the last few years has had tremendous impact upon all aspects of business and government. From its beginnings in the early 1950s with the introduction of the first general purpose electronic digital computers, the data processing environment has expanded and has become more and more complex, impacting upon more and more individuals within an organization, as well as the environment the organization operates in. The tremendous technological breakthroughs in computer hardware has led to increased availability and use of computers. New concepts had to be developed in order to more effectively understand and manage the data processing environment. Once management was able to recognize and describe this complex environment, it developed more sophisticated tools and techniques to deal with this environment.

Information Resource Management (IRM) has become the watchword of the eighties in regards to data processing. With the automation of the functions of an organization, data processing becomes vital to that organization. The information, and the data from which it is produced, become critical resources which any organization must manage efficiently and effectively. This management can be implemented through establishment of a database administration function highly placed in the organizational hierarchy. The database administrator is responsible for the

entire database environment of an organization, and must enforce effective and efficient administration of data resources and compliance with promulgated regulations by organizational personnel. In an effort to control the entire database environment, the administrator should make use of available software tools, among them data dictionary systems and database management sysems.

A data dictionary system provides effective centralized control of data resources in a uniform manner across organizational boundaries. Data dictionary systems and database management systems complement one another in the management of the database environment. A data dictionary is vital in the effective collection, specification and management of the total data resources of an organization.

This thesis will explore the role of data dictionary systems in IRM. In order to better comprehend this role, key concepts of today's complex data processing environment will be discussed. Included in this discussion is the evolution of information systems and database concepts from their earliest precursors, the importance of software life cycle management and the implementation of IRM. A key component in effecting IRM is the data dictionary, which is covered in detail from its evolution to its present functions and future directions. Finally, the effect of legislative action upon the implementation of IRM and the use of data dictionaries by Federal agencies, particularly the United States Navy, will be explored.

II. CONCEPTS

A. INFORMATION SYSTEMS

The first general purpose electronic digital computer was introduced in 1951, inaugurating the "Computer Revolution" or "Information Revolution" of the twentieth century. Improvements in computer hardware and associated developments in software led to the introduction of subsequent generations of machines, the major characteristics of which are detailed if Fig. 1. Improvements in hardware which produced faster, smaller, cheaper machines capable of processing and storing greater amounts of data have led to the explosive proliferation of computer usage into every facet of business and government.

Early generations of machines were mainly focused upon hardware due to its cost. Software applications were initially a minor consideration. This focus has shifted over the years due to the dramatic decrease in the price of hardware and the equally dramatic increase in memory capacity, as can be seen in Fig. 2. Software became an important factor in effective and efficient utilization of the vast data processing and storage capacity of later generation machines. The change in the relative cost of software as opposed to hardware since the introduction of the first generation machines is shown in Fig. 3.

Computer systems, composed of hardware and associated software, require data for processing. Without this data there is

MAJOR CHARACTERISTICS	FIRST GENERATION	SECOND GENERATION	THIRD GENERATION	FOURTH GENERATION
YEAR INTRODUCED	1951	1959	1964	early 1970s
ELECTRONIC CIRCUITRY	Vacuum tubes	Transistors	Integrated semiconductors circuits	Large scale integraged (LSI) semiconductor circuits
MAIN MEMORY	Magnetic drum	Magnetic core	Magnetic core	LSI semiconductor circuits
SECONDARY MEMORY	Magnetic tape Magnetic drum	Magnetic tape Magnetic disk	Magnetic disk Magnetic tape	Magnetic disk Floppy disk Magnetic bubble
INPUT MEDIA/ METHOD	Punched cards Paper tape	Punched cards	Key-to tape disk	Keyboard/video data entry Optical recognition

Figure 1--Characteristics of Computer Generations [Ref. 1]

OUTPUT MEDIA/ METHOD	Punched cards Printed reports	Punched cards Printed reports	Printed reports Video display	Video display Audio response Printed reports
SOFTWARE	User written programs Machine language	Packaged programs Symbolic languages	Operating systems High-level languages	Data base management systems (DBMS) User-oriented languages
OTHER CHARACTERISTICS	Batch processing	Overlapped processing Real time processing Data communications	Time sharing Multi-programming Multi-processing Minicomputers	Micro-programming Virtual memory Distributed processing Word processing Microcomputers

Figure 1(continued)--Characteristics of Computer Generations

TREND IN COMPUTATION SPEED
(in multiplications per second--MPS)

First generation -------------------- 300 MPS
Second generation ------------------- 200,000 MPS
Third generation -------------------- 2 million MPS
Fourth generation ------------------- 20 million MPS

TREND IN COMPUTATION COST

Average cost of doing 100,000 multiplications
1952 = $1.26 1958 = 26¢ 1964 = 12¢ 1974 = 1¢
Today the cost is a fraction of a cent

TREND IN SPEED OF AN ELECTRONIC LOGIC CIRCUIT

Mid 1950s (vacuum tube circuit) = 1 microsecond
Early 1960s (transistorized printed circuit = 100 nanoseconds
Late 1970s (integrated circuit chip) = 5 nanoseconds
Mid 1980s (integrated circuit chip) = 1 nanosecond?

TREND IN CIRCUIT COST

Per integrated circuit chip
1964 = $16 1972 = 75¢ 1977 = 15¢ 1985 = 1¢ ?
Per bit of integrated circuit memory chip
1973 = 0.5¢ 1977 = 0.1¢ 1985 = 0.005¢ ?

TREND IN RELIABILITY OF ELECTRONIC LOGIC CIRCUIT

Vacuum tube = one failure every few hours
Transistor = 1000 times more reliable than vacuum tube
Integrated circuit = 1000 times more reliable than transistor

Figure 2--Costs and Performances of Electronics [Ref. 2]

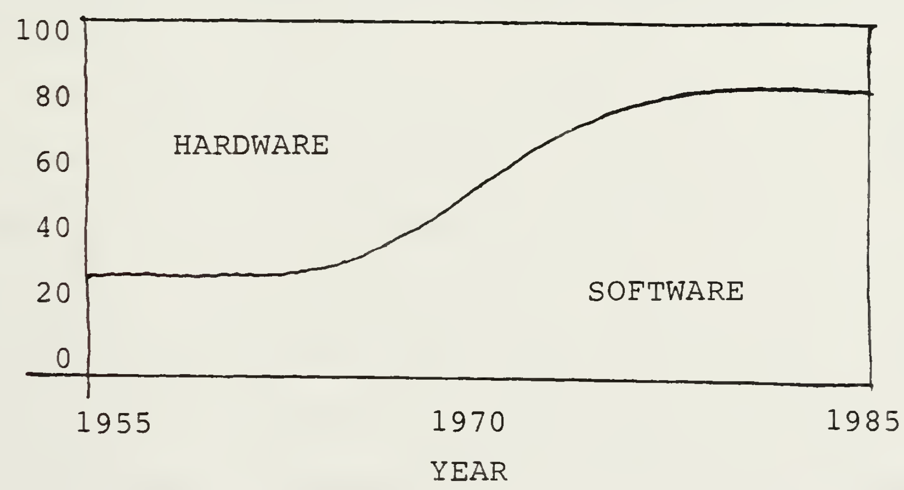

Figure 3--Hardware and Software Cost Trends [Ref. 3]

no reason for the system to exist. Additionally, personnel are required to operate the computer system which processes this data. Finally, these personnel should have prescribed procedures for effective and efficient operation of the system. Hardware, software, data, personnel and procedures are, therefore, mandatory components of an Information System (Fig. 4).

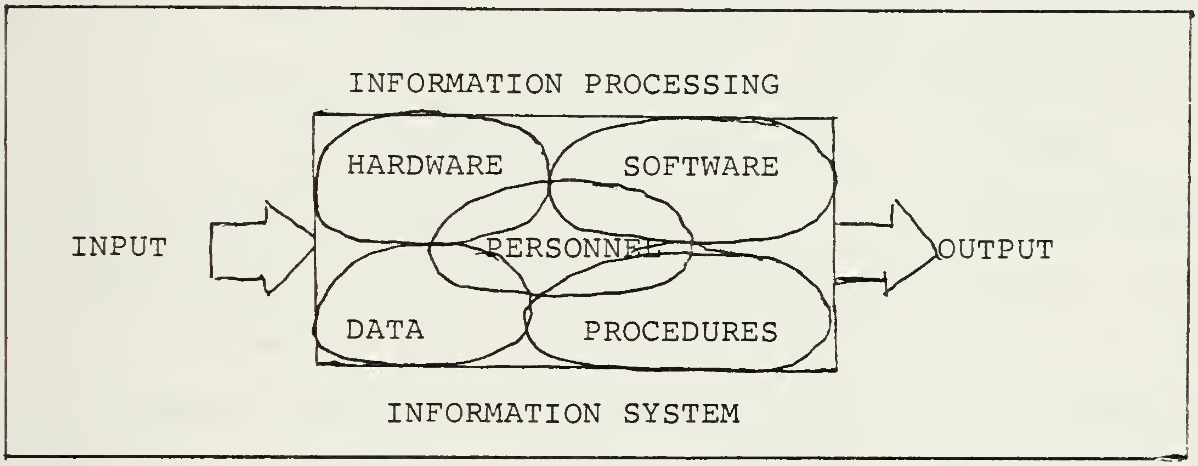

Figure 4--An Information System [Ref. 4]

An Information System may be defined as:

> a system which uses personnel, operating procedures, and data processing subsystems to collect and process data and disseminate information in an organization. [Ref. 5]

It is no longer possible to consider only the hardware facet in data processing. Sophistication has led to the need to consider every facet of an Information System and their interaction with each other. Of primary concern is the rising cost of personnel and projected manning shortfalls in the next twenty years, which increases the demand by an organization for the most effective and efficient management of an Information System. One

method for this improvement is increasing the number and/or use of software tools and implementing improved management procedures/techniques.

B. DATABASE

Originally, computers processed programs composed of data and instructions to produce the information desired by an organization. Due to limitations of memory capacity and the awkwardness of coding in machine language, early applications were usually limited to automation of repetitive daily operations. Second and third generation machines, with their increased memory capacity and more efficient data processing software innovations, allowed for more and more of an organization's operations to be computerized, but still remained relatively oriented toward automating the paperwork of an organization. Each application was developed independently, viewing its associated data in a proprietary fashion, creating and maintaining them as needed.

The development of third and fourth generation machines gave the user increasingly extensive processing capabilities, but required a more comprehensive view by the user in order to fully realize their potential. In the very early days of computing, data and instructions were intermixed in the program. The most complex data structure applicable at this time was a "field" (or "string") consisting of meaningful groups of single alphanumeric or other symbolic "characters". More complex data structures evolved--the "record" composed of related fields and the

"file", itself a group of related records (Fig. 5). File processing, which was utilized by second and third generation

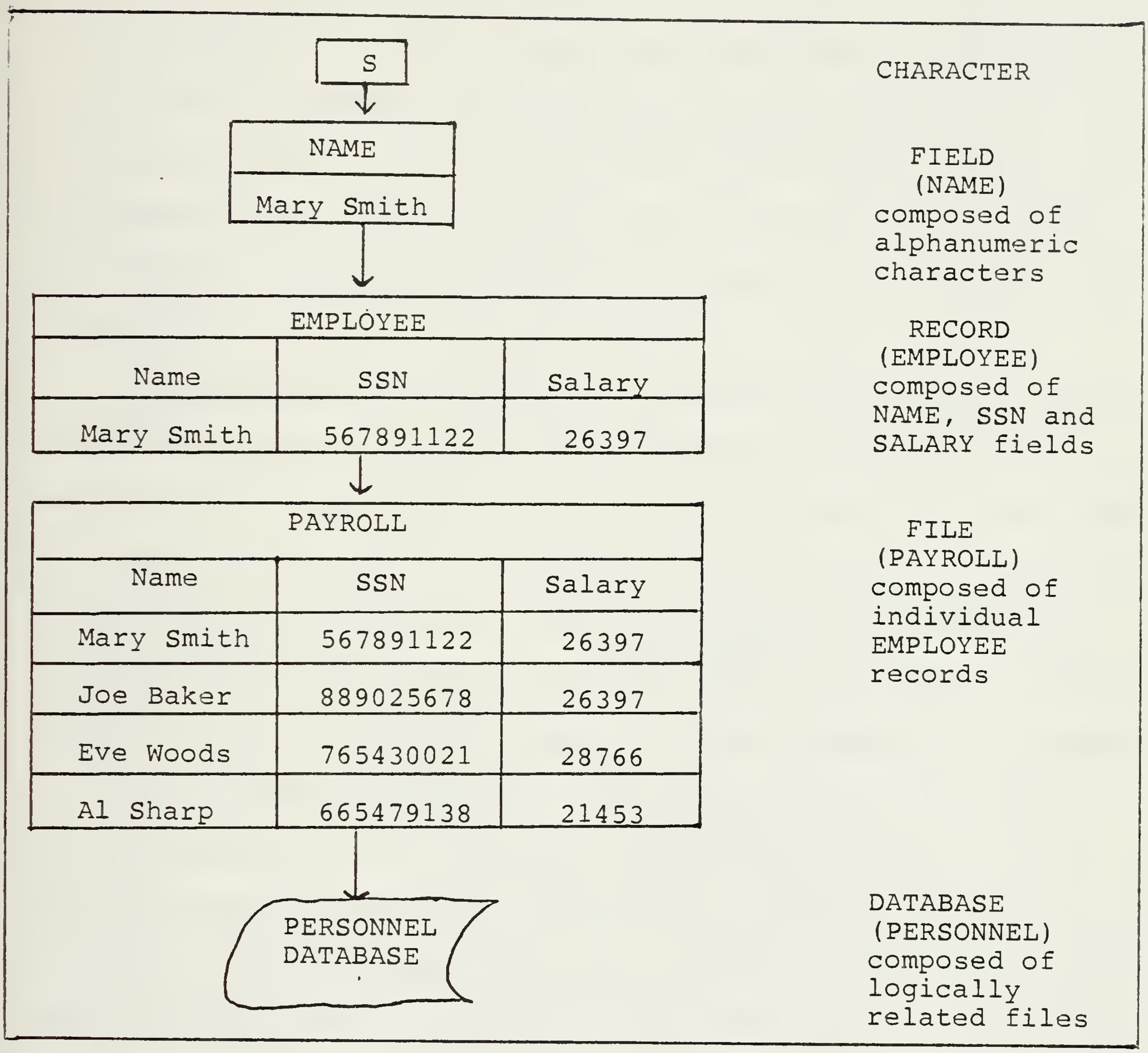

Figure 5--Hierarchy of Data Elements

machines, does not allow flexibility in data processing. Each application maintains its own files and reocrds separate from other applications, making data dependent upon the application

which utilized it. This leads to inconsistency and incompatibility in the data, especially when data has been updated. An additional problem was the absence of an ability to combine data from separate files and records quickly and easily. New applications had to develop data files from scratch, increasing the development cost. One-of-a-kind applications were usually not implemented due to the cost and time required to develop them. The database concept was developed in order to solve these problems (Fig. 6).

A database is a nonredundant collection of logically related files. By definition, data redundancy prevalent with file processing is eliminated. Data is held collectively. More than one application may utilize the same data, allowing independence of data from applications. More sophisticated programming is possible, and new applications can be quickly and easily implemented. One-of-a-kind applications become economically feasible.

While database processing solves many of the problems of file processing, there are some disadvantages attached to its use. Software programs are required for database management (i.e., Database Management Systems, or DBMS), which can be expensive to purchase or develop. This additional software often requires increased hardware resources. The additional software interface increases the processing time, thereby increasing the cost of data processing. Database processing is complex, requiring more sophisticated programming and more highly qualified

operating personnel. Recovery and backup, in case of failure, is more difficult than with simpler file processing systems.

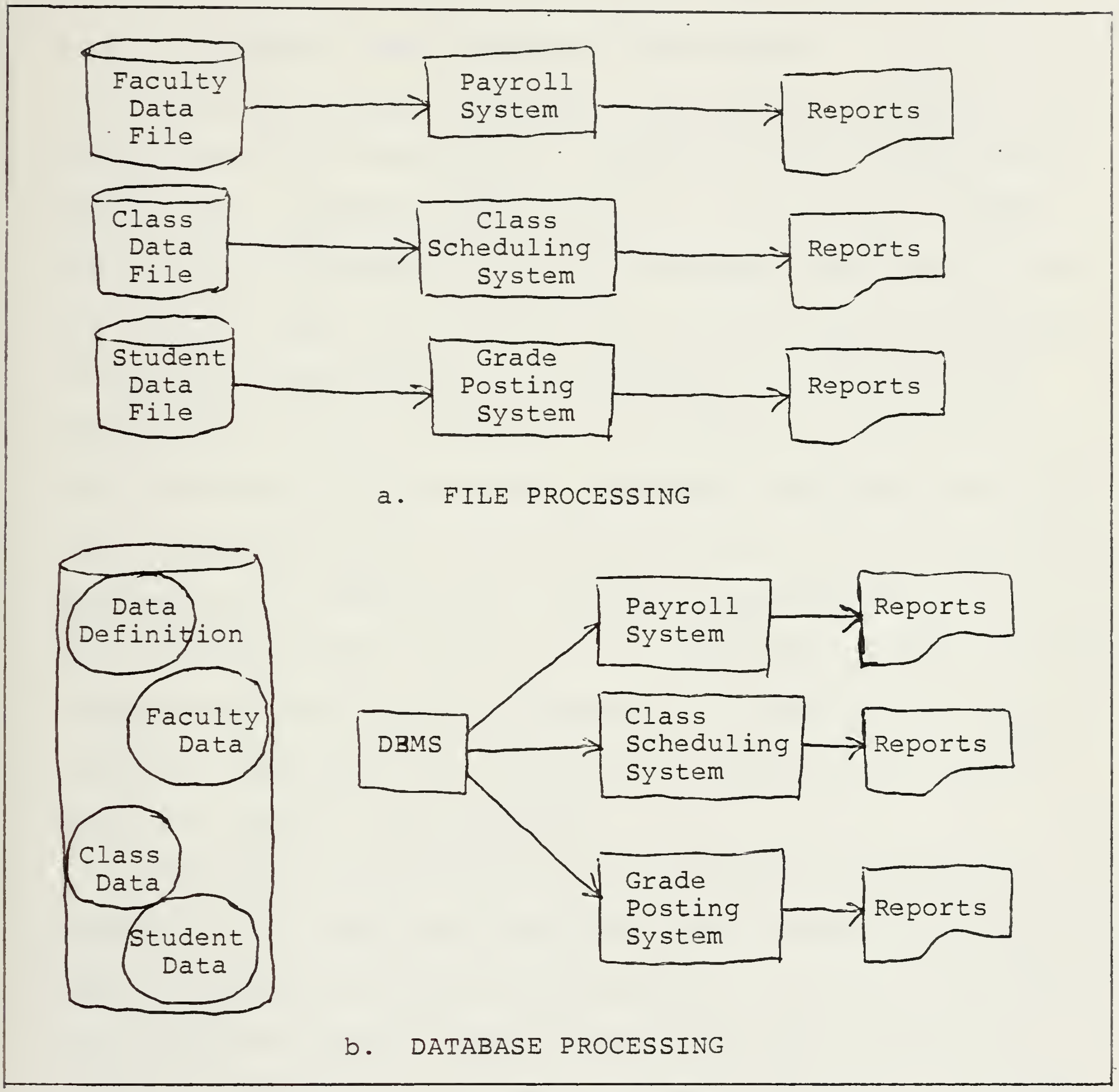

Figure 6--File/Database Processing [Ref. 6]

Finally, database processing has increased vulnerability to failure. However, even with these considerable drawbacks, the advantages of using database processing make it extremely desirable in today's data processing environment.

In order to design a database, one requires conceptual representations of the real world. The most basic is the entity, which is the conceptual representation of an object. Kroenke [Ref. 7] defines an object to be a phenomenon which can be represented by a noun. During the design of the logical database, entities are unrestricted by the constraints of the computer. Entities will not be transformed into computer record format until the design of the physical database. Entities have attributes which characterize and describe them. In the real world, objects may be related to one another in associations. The conceptual representation of this is a relationship between entities. Relationships may also have attributes, as entities do. The conceptual representations pertaining to data must be defined during the design of the database.

A database is a self-describing collection of integrated files [Ref. 8]. The self-describing aspect refers to the fact that the database contains a description of its own structure. The integration aspect refers to the fact that a database is not just a collection of files, but also contains the relationships which exist among these files. In order to process the database, one or more keys are necessary. The key is a field

which identifies a record. A key may be unique, identifying only one record, or nonunique, identifying a group of records. Database processing may also utilize record relationships.

A database has three views of data: schema, subschema and physical. The schema, or conceptual view, is the complete, logical view of all the data in the database. From the control standpoint, however, it is inadvisable to allow every application access to the entire database. The subschema, or external view, defines a subset of the schema which is accessed by a specific application. Since different applications may require the same data to some extent, subschemas may overlap. Subschemas may also reorganize the schema, depending upon the capabilities of the DBMS used. The third view, the internal, or physical view, describes how the data is physically arranged and how it is allocated to files. Each of these views must be defined before database processing can occur. Usually, management personnel define schemas and subschemas, while the DBMS defines the internal view when the database is defined. The variety of views of the same data which database processing offers allows subschemas to be tailored for the needs of the specific application. This means that each user sees the data in a familiar and useful format, even though the data is centralized and shared.

C. SOFTWARE LIFE CYCLE

Large, complex software systems require a large amount of effort and time to develop and are in use for an even longer

time. A number of distinct stages in the entire life span of the software can be identified. They are components of the software life cycle. These stages are:

(1) Specification
The software requirements (i.e., the system functions and operational constraints) must be established and specified.

(2) Design
A software design must be derived from an analysis of the software requirements.

(3) Implementation
The software design must be converted into a programming language which can be executed on the target computer.

(4) Testing
The implementation must be tested to ensure that the completed system meets the software requirements.

(5) Operation and Maintenance
The system must be installed and used. If system errors are discovered, these must be corrected and changes to the original requirements may involve adding to the system.

Software development, which encompasses the first four stages of the software life cycle, is an iterative process with feedback from each stage to previous stages. During development, requirements may be clarified, implementation may reveal design flaws, testing may reveal errors in preceding stages and operation may reveal errors which were undetected at earlier stages. In each instance, a change to correct a detected error will involve a recycling through the applicable stages of the life cycle.

The operation and maintenance stage of the life cycle accounts for the major portion of the total software cost. Typically,

operation and maintenance costs are four times as much as the total cost of software development (stages one through four), but can be as high as fifty times the cost of software development [Ref. 9]. Therefore, efforts aimed at reducing total life cycle costs are best concentrated on reducing the costs of the maintenance stage. As maintenance requires modification of the software, maintenance costs can be reduced by ensuring that the software is understandable and easy to change. This implies that specifications must be unambiguous, design and implementation tailored so that the software is composed of easy to modify modules and validation techniques are used to minimize the number of undetected errors. The earlier in the life cycle these errors are detected, the easier and less expensive they are to correct.

In order to perform effective validation of the life cycle stages, requirements and design specifications must be unambiguous. Unambiguous specifications are produced when formal notations are used and these specifications can be checked using software tools which have been developed for this purpose. Each stage should be thoroughly and properly documented, and where feasible, automatically checked for consistency and completeness. While this may increase the costs of software development, this increase is more than compensated for by a reduction in maintenance costs and, therefore, results in a reduction in the total software life cycle cost.

D. INFORMATION RESOURCE MANAGEMENT

1. Definition

Due to the proliferation of computers, the increasing complexity and variety of applications and the scarcity of highly qualified personnel at ever escalating salaries, management became aware of the increasing need to manage Information Systems effectively and efficiently to the benefit of the organization. The Information Resource Management (IRM) concept resulted from this recognition of management for the need to treat information as it would any other resource critical to the organization. The Workshop on Data Dictionary Systems and Information Resource Management (1980) defined IRM as:

> whatever policy, action or procedure concerning information (both automated and non-automated) which management establishes that serve the overall current and future needs of the enterprise. Such policies, etc., would include considerations of availability, timeliness, accuracy, integrity, privacy, security, auditability, ownership, use and cost-effectiveness. [Ref. 10]

Therefore, information policies, actions and procedures must be planned and executed organization-wide in order to truly treat information as a critical organizational resource. IRM is a reflection of the shift from systems designed around the processing methodology to systems designed around the data to be processed.

2. Database Administrator

The recognition of the IRM concept by managers leads to the recognition of the need for disciplined control of the

data of an organization. This control is incorporated in a set of management procedures and technical functions which comprise the emerging disciplines of database administration.

Database administration encompasses all the technical and management activities required for organizing, maintaining and directing the database environment, which is considered to consist of the following:

-- the database (including automated and non-automated data)

-- the database administrator (DBA) who manages the database environment

-- the software tools used in data administration and processing

-- the users of the database

The basic functions performed by the DBA include database definition/redefinition, selection and procurement of hardware/software/services, database design/redesign, database creation, database security/integrity, database maintenance/management, database performance monitoring and evaluation, database enforcement, liaison with users/staff/management, and conversion of non-database systems to database systems [Ref. 11].

The DBA is responsible for planning, design, development, implementation, testing, documentation, operation and maintenance of the entire database environment. Due to the impact of database administration upon the entire organization, the DBA position should be placed high in the organizational hierarchy to ensure its success in enforcing effective and

efficient administration of data resources and compliance by members of the organization with database rules and regulations.

3. Software Tools

a. Database Management Systems

A DBMS is a software tool which provides an integrated source of data for multiple users, while presenting different views of the data to different users. It can be characterized as generalized software which provides single flexible facility for accomodating different data files and operations while demanding less programming effort than conventional programming languages. It features easy access to the data, facilities for storage and maintenance of large volumes of data, and, most importantly, the capability for sharing the data resources among different types of users. Since DDS are concerned with the management of data elements, it is logical that a strong relationship between DDS and DBMS exists. Some DDS interface with a variety of DBMS, while others require a specific DBMS in order to operate, while still others are embedded in an existing DBMS.

DBMS evolved from the attempt to develop integrated application systems which were complicated by intricate data structures. The earliest DBMS was developed in the 1960s, based on hierarchic, network and inverted-tree data models. Continuing improvements in DBMS have caused these early models to be mostly superseded. Fig. 7 depicts the relationship of

the currently popular models. The database model is a vocabulary for describing the structure and processing of a database, and can be used for both logical and physical design of the database. It is also the basis for categorization of DBMS products. The database model is composed of Data Definition Language (DDL), which is used to define the structure of the database, and Data Manipulation Language (DML), which is used to describe the processing of the database.

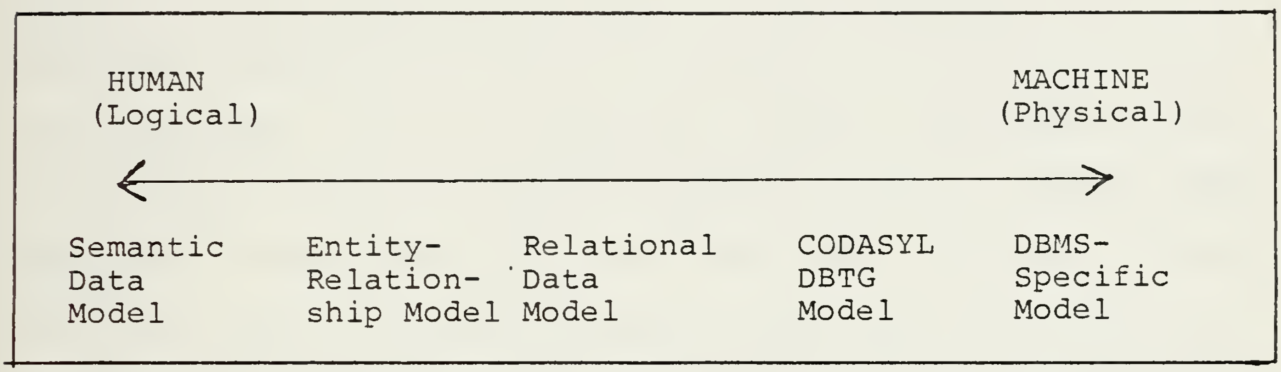

Figure 7--Relationship of Data Models [Ref. 12]

While Fig. 7 depicts five data models, only three of these have actual DBMS products available. The Semantic Data Model (SDM) provides a means for expressing meaning as well as structure of database data. As such, it is the most logically and least physically oriented model, and lends itself particularly well to logical database design. The Entity-Relationship (E-R) Model is primarily a logical database model with some physical aspects. Though these models are convenient and lend themselves to the logical design of databases to describe

what the user wants to see, neither model has a DBMS implementation at present and must be translated into physical database constructs once a particular DBMS product has been selected.

The Conference on Data System Languages (CODASYL) Data Base Task Group (DBTG) Model is the oldest model listed in Fig. 7. A survivor of the earliest developments in database management, this model was developed in the late 1960s and is a physical database model, providing constructs defining physical characteristics of data, its location, data structures used for implementing record relationships and similar record relationships. Due to its physical nature, the CODASYL model is difficult to use for logical database design. Several DBMS products are available which are based upon this model, however, there are some detracting factors to its use. The model, naturally, is geared towards COBOL, and is not easily implemented in organizations which utilize a language other than COBOL. Also, the model is very complex and somewhat incohesive. Some decisions regarding the model were based on group politics rather than technical merit. Finally, many variants on the core concept create confusion for the user.

The Relational Model was first proposed by Dr. E. F. Codd in 1970, and has been the focus of a great deal of activity, which has been largely theoretical in nature since commercial DBMS products based upon this model have only been available in the last few years. However, this model appears

to be the "new wave" in DBMS implementation. The significance of the Relational Model is that relationships are considered to be implied by data values. The principal advantage of carrying relationships in the data is flexibility. Unlike the CODASYL Model, relationships need not be predefined in the design of the database.

DBMS-Specific Models are those DBMS which do not conform to any of the above models. These DBMS are based upon unique data models and some (e.g., ADABAS, TOTAL, IMAGE) have been commercially successful. Due to the variety among this category of DBMS, it is difficult to establish specific characteristics about them. Fig. 8 gives a brief summary of these models.

The use of DBMS provides significant advantages in reducing the redundancy of stored data, avoiding inconsistencies in stored data, allowing for the sharing of stored data, maintaining data integrity and enforcing standards. However, the actual use of DBMS does not necessarily resolve all problems relating to the data administration function within an organization, especially when the DBMS are used primarily for their storage and retrieval capability. This particular usage is not recommended, but frequently happens anyway. The increasing variety of DBMS products has resulted in instances within an organization where more than one DBMS is employed within that organization, emphasizing the need for a facility which provides

uniform and centralized control of all the data resources of the organization. DDS is a tool which assists the DBA in performing this function.

TYPE	CHARACTERISTICS
SDM	DDL language for storing meaning High level DML No DBMS based on this model
E-R	Entities and relationships modeled as data E-R diagrams graphically show relationships No DML
Relational	Data represented as tables Relationships implied by data Dynamic data relationships Procedural and nonprocedural DML A few DBMS based on this model
CODASYL DETG	Oldest data model Relationship must be predefined Procedural DML Extensive application in industry Many DBMS based on this model
DBMS-Specific	Models vary widely DDI and DML closely conform to features of the DBMS

Figure 8--Summary of Data Models [Ref. 13]

b. Data Dictionary Systems

With the growth of data resources of an organization, effective control cannot be maintained through the use of a DBMS alone. The DDS provides this vital central control function in a uniform manner across boundaries within an

organization. The DDS and DBMS complement one another in the management of the database environment. Many of the benefits realized from the use of a DDS are parallel to those attributed to the use of a DBMS. However, there is a significant difference in that the benefits realized from a DBMS are directly related to the effective computer processing of the data, while the benefits realized from a DDS are directly related to the total data resources of an organization. Since the functions of a DDS coincide with the goals of database administration, it is one of the basic tools utilized by a DBA.

c. Other Software Tools

There are a great number of commercial software products available which can assist the DBA in the management and control of the database. The two most important are DBMS and DDS. Other tools which are useful in database administration may exist as a separate, self-contained piece of software, as part of another piece of software or as a facility or utility of a DBMS, a DDS or an Operating System (O/S). Leong-Hong and Marron [Ref. 14] give the following list:

> Information/Data Retrieval System (IRS)--a program or set of programs which enables the user to retrieve information in a variety of formats; most modern IRS provide interface capabilities, including extensive user-oriented facilities and rapid response to system commands.
>
> Online Query System--a separate program or a DBMS feature which enables the user to interactively obtain information contained in a database.
>
> Data Entry System--provides facilities for automatic data entry and collection. Some provide interactive data entry

facilities, allowing for key verification, limited editing and formatting; others provide batch operation to enable massive data loading.

Editor--facilitates selective modification and correction of data, program and document text. Special purpose editors are available, geared towards entry, modification and editing of data, files, programs or texts.

Flowchart Generator--produces a pictorial diagram of the flow of control and logical paths of a computer program; narrative documentation may also be produced.

Text Processor--a documentation aid that accepts lines of source text interspersed with format control commands, and formats the text into a printable, paginated document with user-designed style.

Report Generator--allows automatic generation of pre-formatted reports on a production basis, or allows definition of ad-hoc reports, via parameters.

Cross-Reference Generator--produces listings of data elements used in files, programs and systems indicating where data elements are being referenced. For programs, it produces listings of the variables (data elements) used in programs, subroutines and systems, indicating where they are being referenced.

Text/File Reformatter--rearranges and structures files according to specifications, and rearranges and structures text and source programs for improved readability.

Data/File Mainenance Programs--perform global changes for all, or selected, records in a file, while reporting changes in data context before and after operations. They may provide data/file edit capabilities, and data items deleted or added may be flagged for audit trail purposes. They may be a separate software package, a utility program provided by an O/S or a feature of a DDS or DBMS.

Data Editing and Validation System--provides the user with the ability to perform data validity test, data editing, error correction and error reporting; or any subset of these tasks.

Data Auditor--examines source data definitions and analyzes data relationships. data structures, formats and storage usages for consistency, validity and efficient utilization.

It may provide a dictionary or catalogue that contains definitions of the data attributes, and characteristics of the data type. It is available as a separate software package, but it is also a feature of a DDS or a DBMS.

Data Security Module--may provide protection over sensitive data by encrypting/decrypting and by controlling access to the sensitive parts of the database. Security can be achieved through encoding/decoding or through execution-time password capabilites.

Test Data Generator--produces test data files, according to specifications, which can be used for testing application software.

Optimizers--apply changes directly to program source code in order to make them run more efficiently by reducing run-time or core requirements. They may perform analysis of the program for undetected errors and optimal logical flow.

Automatic Space Generators--find available space for programs or files that are awaiting processing.

Scheduler--allocates available computing resources in order to optimize the use of resources to daily workloads. They may produce reports indicating the areas where optimization of the resources may occur.

Project Manager--provides data collection, storage, and reporting facilities aimed at personnel time and task accounting. They may be coupled with other productivity and scheduling management aids.

Librarian--facilitates organized and economical storage of programs, texts, data sets, and object modules for centralized retrieval and updates. They may collect accounting data to assist in storage allocation.

E. SUMMARY

The explosive proliferation of computers has led to the increasing importance of developing and implementing various management concepts for effective and efficient operation and control. Emphasis has shifted from viewing a single component, computer

hardware, to consideration of the entire information system, composed of hardware, software, data, personnel and procedures. Increasing complexities in storage and retrieval of data has led to the development of the database concept. The critical role information has come to play in an organization's success has led to its recognition with the IRM concept. Finally, the increasing importance and visibility of the person or persons in charge of the organization's data has increased the interest in and need for effective and efficient management and control of data. One software tool which can provide this management and control is the DDS.

III. DATA DICTIONARY SYSTEMS

A. EVOLUTION

Data Dictionary Systems (DDS) are a relatively recent innovation in the field of data processing (DP). The earliest commercially available products were introduced in the early 1970s, but these systems were relatively primitive and provided only a few functions. Impetus has gathered for improvement and extension of DDS due to management acceptance of the IRM and Software Life Cycle Management concepts, as well as the ever increasing complexity of an organization's database. Control of the data resources is vital to the future success of an organization, and this concern for control is evident in the increasing number of organizations implementing DBA functions within the organization and utilizing software tools for management and control.

Initially, the DP environment was relatively simple, requiring little in the way of management beyond coding programs and scheduling jobs to be run. Management of hardware resources was of primary concern to a relatively small DP department which was located several layers down in the organizational hierarchy. This was a reflection of the initial automating of daily repetitive functions. With the growth and improvement of computer systems and evolution into a more complex information system concept, there was a concomitant demand for more effective and

efficient management. As early, simple hardware management and scheduling techniques were automated through the use of more sophisticated software (i.e., O/S), management could concern itself with more complex data management, integration and control tasks.

Extensive DP capacity led to the development of the database concept. Once management embraced this concept, it was necessary to develop techniques and software tools to manage it, resulting in DBMS being introduced in the early 1960s. With the growth in complexity and amount of data an organization required for operation, simple manual listings of the contents of data records, files and even databases was ineffective and outmoded. While DBMS was an effective tool for storing and retrieving data in a database and provided some control, it did not provide enough control to meet the objectives of IRM. DBMS reflects the emphasis prevalent in the late 1960s to early 1970s on data and data management. The emphasis has shifted in the 1980s to information and IRM, which requires more than just data management in order to be effective.

Since DBMS were developed before DDS, there is a natural tendency to view DDS as subordinate to DBMS, especially in instances where a DDS-like function is part of the DBMS or where DDS is dependent upon DBMS for operation. However, the increasing interest and improvements in DDS, and the development of DBMS-independent DDS, has caused it to evolve into a complex

software tool which should be considered equal to and allied with DBMS to effect IRM.

The British Computer Society (BCS) established a study group, the Data Dictionary System Working Party (DDSWP) in January 1975. Over a period of time the BCSDDSWP worked to produce a report on the need for and the facilities which should be provided by a DDS and related database design and operational aids. They studied the then currently available DDS and related design aids, identified data recording and analysis needs for the design of information systems, specified requirements for aids to database design, maintenance and operation, and considered which of these requirements were automatable. The report of this study group was published in late 1977. This study emphasizes the shift which began in the mid 1970s to expanding the functions of a DDS from a software tool which was primarily involved with cataloging the data in an existing database into an adjunct to designing the database itself. Utilizing DDS in design of new software systems would assist in more effective management of the software life cycle and assist analysts and designers in determining undetected errors early in the software life cycle. It would also make maintenance of software easier and cheaper.

B. DEFINITIONS

Definitions of DDS range from Cardenas' (1979) relatively simplistic "centralized repository of data about data" [Ref. 15]

to the National Bureau of Standards' (NBS) (1982) definition of a DDS as a resource manager which is:

> an integrated repository that provides data necessary for managing data, where data management includes the planning, control, direction and organization of data. [Ref. 16]

Other definitions include those of the BCSDDSWP (1977):

> a tool for recording data and processing information about the structure and usage of data [Ref. 17];

Leong-Hong and Marron (1977):

> a software tool that provides the means for defining and describing the characteristics of a database, as opposed to the contents of a database [Ref. 18];

and Allen, Loomis and Mannino (1982):

> an automated information system which achieves control of data by providing an inventory of data, control of costs of developing and maintaining applications by providing accurate and complete data definitions, and independence of metadata (i.e., data about data) across computing environments, improving resiliency to hardware and software changes [Ref. 19].

The variety of definitions gives some idea of the evolving scope and increasing complexity of DDS. Originally envisioned as a database management tool, separate from and subsidiary to DBMS, DDS has evolved into being a primary componenet of a system for information management. In fact, it is proposed that DDS is an outdated concept and too restrictive for the IRM concept, which requires an Information Resource Dictionary System (IRDS). An IRDS is:

> an information system with automated support which documents the information environment of an enterprise, supports the operational aspects of that information environment, illustrates the interrelationships of information environment

components, and documents the locations of all components of the information environment. The Information Resource Dictionary (IRD) is the actual database manipulated by the IRDS software [Ref. 20].

Due to the increasing interest in the rapidly evolving nature of this field in recent years, terminology is somewhat confusing. One author speaks of a DDS while another refers to data element dictionary/directory system (DED/DS) and, of course, the most recent developments are towards IRM and IRDS. NBS [Ref. 21] defines the following terms:

Data Catalog: a listing of data elements

Data Element Dictionary: describes each data element

Data Element Directory: locates each data element

Data Element Dictionary/Directory: describes and locates as well as lists each data element.

Plagman [Ref. 22] further elucidates the difference between dictionary and directory functions by the type of user: dictionary users are human, while directory users are (usually) systems components (i.e., hardware/software). Since most commercially available packages have both dictionary and directory functions, apparently even those authors who speak of DDS are actually referring to a DED/DS, which only adds to the confusion. In this thesis references to DDS will imply that a directory function is also available unless specifically stated otherwise.

C. FUNCTIONS

In addition to the confusion generated by differences in terminology, the broad divergency of opinion as to the scope of a DDS further clouds the picture. The scope of a DDS can be quite narrow, covering only the database definitions necessary to support a DBMS, or exceedingly broad and grandiose, covering all data important to an organization, as with the IRM concept. The early DDS, and many installations' initial usage, centered around the directory functions, i.e., the DDS became the main database definition interface. This initial and basic function, the capture and documentation of data elements, their definitions, some of their descriptive attributes and some logical grouping of these elements, has remained relatively constant over the years.

In this aspect, a DDS must be able to identify data elements which are synonyms, homonyms or aliases. Synonyms are different names for the same data element which have become accepted due to common organizational usage. Homonyms are the same name having different meanings according to the context in which they are used. Aliases are different names for the same data element which are determined for DP technical reasons. These may be planned, as in the case of different programming languages, or unplanned, as in the case where no standards exist. In some situations, identifying occurrences of synonyms, homonyms and aliases and relating them to a single naming scheme is a large,

difficult and, occasionally, impossible task to carry out. It can be seen that it is not so much a situation of a DDS allowing an organization to standardize this aspect of DP as it is a necessity to have standards in order to facilitate optimal DDS function. However, a DDS permits more information to be recorded about elements, records, databases, etc., than what is available with just a database definition facility for a DBMS. The ability to document information about reports, users, programs, etc., has generated the impetus to develop DDS into a software documentation tool to support effective and efficient software life cycle management.

Allen, et. al. [Ref. 23] delineates the components of a DDS as a database of metadata (i.e., data describing data, processes, users and processors of an organization) retrieval and analysis capabilities, management tools and functional interfaces. The metadata can be represented by a data model composed of entities, relationships and attributes, equivalent to the concepts used in DBMS. Various attributes which can be used for an entity, or a relationship, in a DDS include type, range, length, unit of measure, usage, language names, repetitions, keys, defaults and display formats. Relationships between entities of a typical DDS are illustrated in Fig. 9.

The typical functions performed by a commercially available DDS are displayed in Fig. 10. These functions are:

(1) Database Maintenance:
interprets and processes requests to add, change or delete contents of the database.

(2) Extensibility:
enables the database structure to be extended by the definition of additional entities, relationships and attributes.

(3) Report Processor:
provides predefined reports, the ability to customize reports and user defined reporting capability. Common predefined report types include:
   (a) name listings
   (b) relationship reports
   (c) detail reports
   (d) summary reports
   (e) matrix reports
   (f) graphical reports

(4) Query Processor:
allows English-like queries of the database (used for low volume retrievals).

(5) Convert Functions
reads application programs, libraries, and schemata and generates input transactions for the Database Maintenance Function (above) which describe the detected metadata.

(6) Software Interface:
provides a formatted pathway enabling DDS to provide metadata to other software systems and enables these systems to retrieve and update information in the database.

(7) Exit Facility:
enables the vendor-supplied routines to be extended (not available in all DDS).

(8) Database Management:
performs database management tasks, e.g., security, integrity, concurrency control, and internal access for the database. This function is often performed by utilizing an existing DBMS, however, DBMS generally does not provide all available subfunctions of this function.

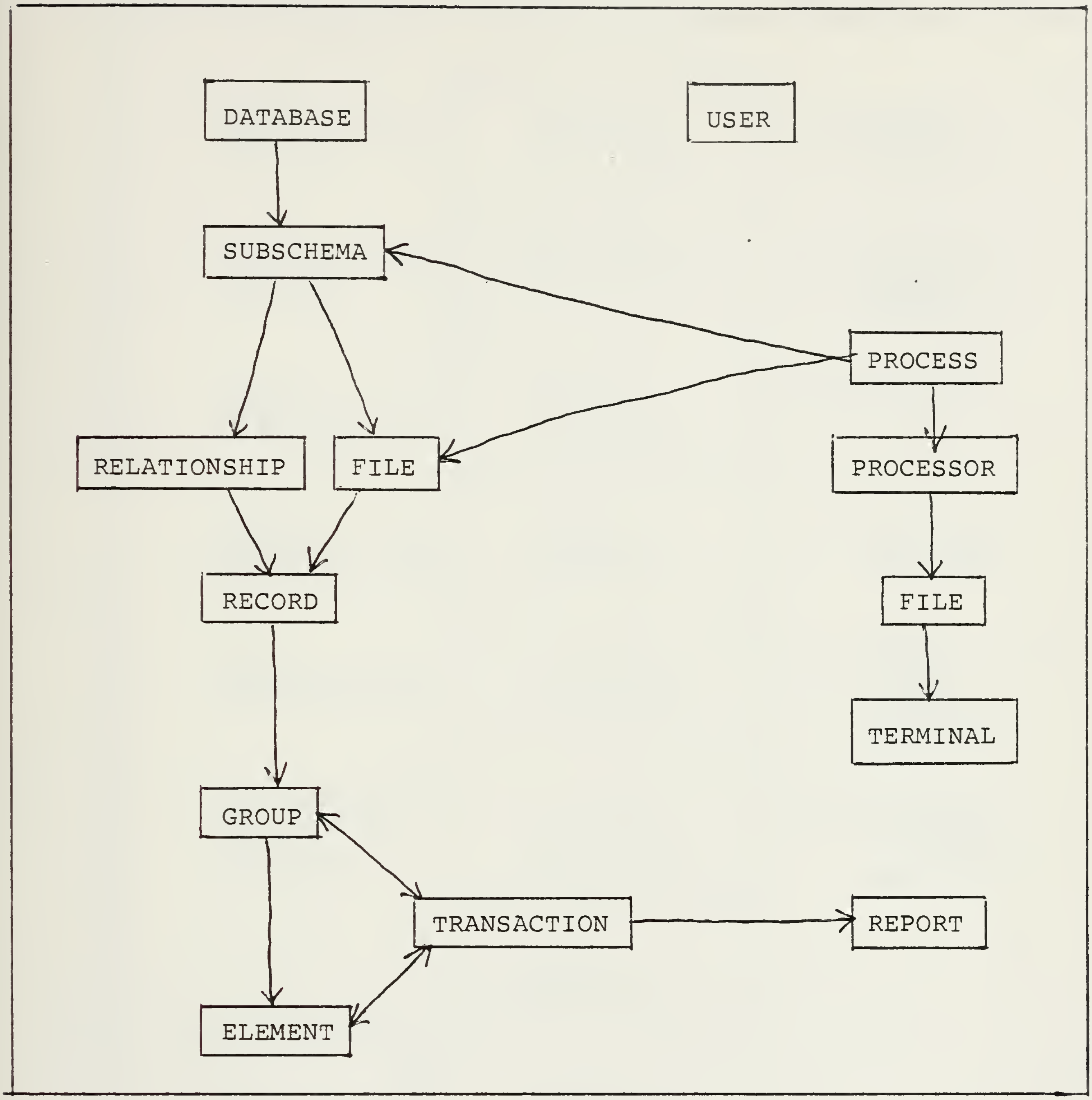

Figure 9--Logical Structure of a Typical DDS [Ref. 24]

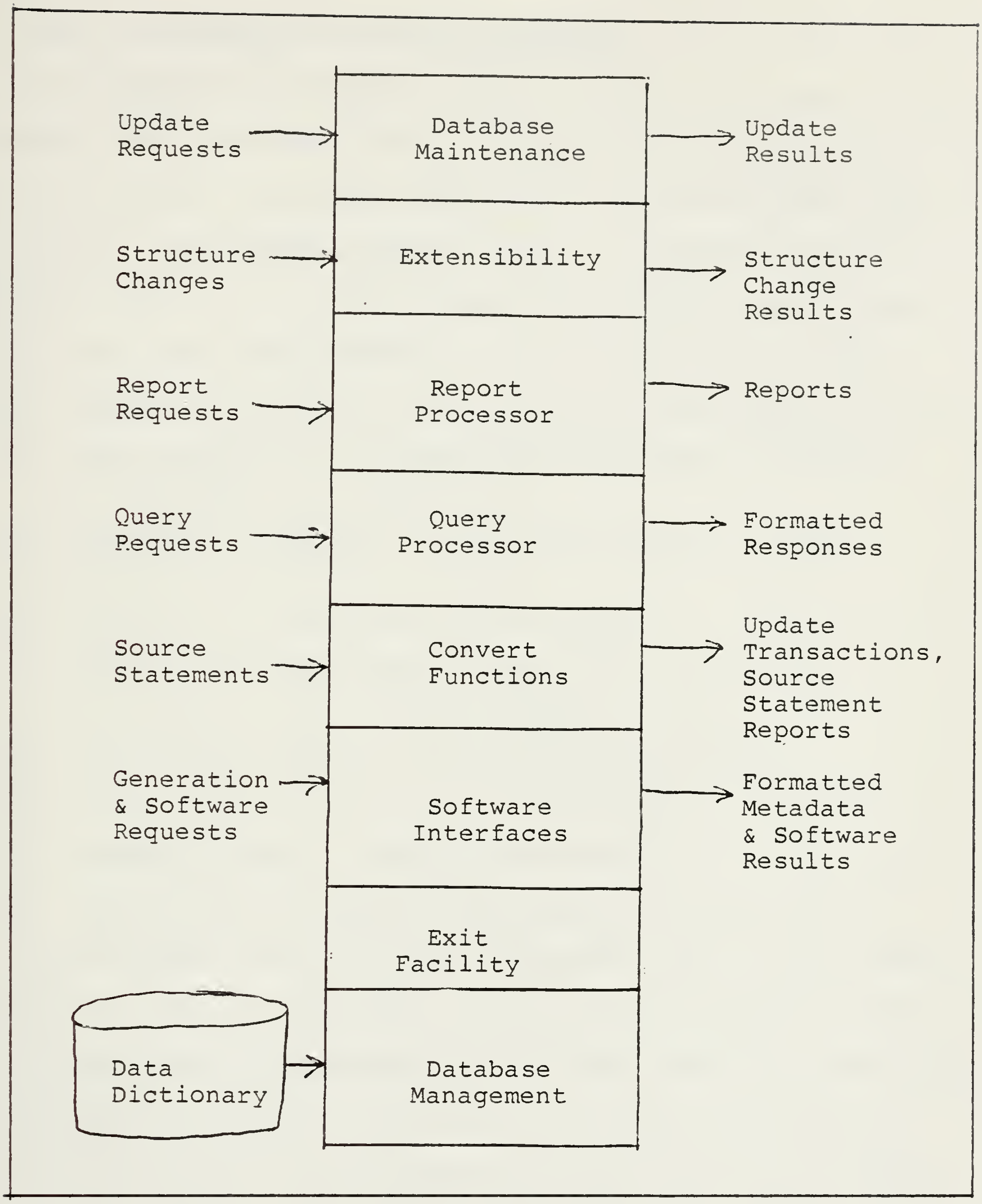

Figure 10--DDS Functions [Ref. 25]

D. ROLE IN SOFTWARE LIFE CYCLE MANAGEMENT

The software documentation feature was not an aspect of the original DDS, which were more in the line of automated lists of data elements, but became the goal of developing DDS in the mid 1970s. The BCSDDSWP Report [Ref. 26], published in 1977, advocated the use of a DDS throughout the complete specification, design and implementation stages of the software life cycle. Particular functions which could be performed would be:

(1) Data analysis, to determine the fundamental structure of the data of an organization

(2) Functional analysis, to determine the way in which events and functions use data

(3) Database or conventional file design

(4) Storage structure design, where this is a further refinement of the initial database or file design

(5) Operational running of the application systems

(6) Collection and evaluation of performance statistics

(7) Database tuning to improve performance

(8) Application maintenance and database restructuring

The BCSDDSWP Report further recommended that the DDS should provide two sets of facilities. One set would record and analyze requirements independently of how they were to be met, the "conceptual data model". The second set would record design decisions in terms of the database or file structures implemented and the programs that would access them, the "implementation data structure". For both facilities it is necessary to record data usage as well as data structure, giving rise to four

areas of information which can be identified. Fig. 11 depicts these four areas.

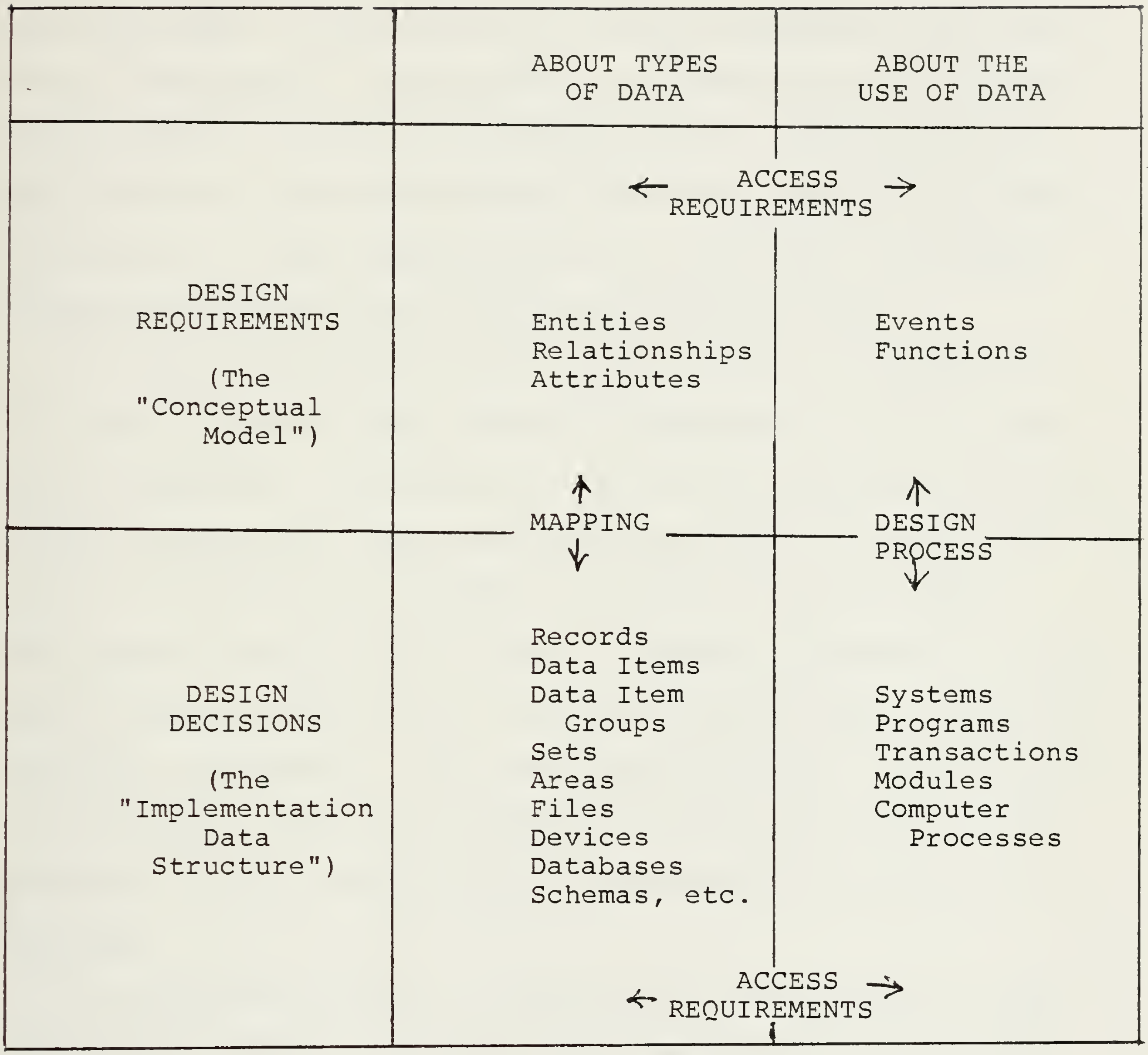

Figure 11--The Information in a Data Dictionary [Ref. 27]

The DDS should relate definitions of the implementation data structure to the parts of the conceptual data model they describe

(i.e., records and items to entities). Recording this mapping documents the design decisions and clarifies the decisions which have been made. There should be a single conceptual data model for an organization, containing all new definitions in addition to the updated versions of old definitions. However, due to the evolution of data structures over time, there will be several versions of each implementation data structure which operationally must have clearly defined changeover times.

In the specification stage of the software life cylce, use of a DDS will assist the analyst in recording the flow of data across functions. Additionally, conflicting usages can be identified and resolved, and redundant data removed from the database or procedures implemented to ensure consistent update. The analyst can also use the DDS to predict the impact of a proposed change and define what actions should be taken to prevent unwanted side-effects. For the analyst the DDS is a device for collecting all the facts necessary for the clear and complete statement of the problem and for providing data to test the solution.

In the design stage, the designer has the conceptual data model for a global view of the organization and where the new system fits within it. Verification and validation of specifications should be completed, as well as determining impacts upon present systems, if any. During this stage an implementation data structure is constructed. The DDS provides

flexibility by allowing individual findings or decisions to be recorded in the appropriate places in the dictionary and provide reference to any item of data of interest to the individual. The DDS not only provides a guide to the project under consideration, but also to the progress and to the documentation. It is much easier to update a DDS than any manual system and, therefore, the DDS provides the most current and reliable form of cross-referencing system available for use in the softwar life cycle.

The implementation stage is more dependent upon hardware and supporting software than are the specification and design stages. If the conceptual data model and functional description are developed in parallel, consistency is ensured. The implementation data structure and programs can then be designed with reference to conceptual/functional model and implementation constraints. The DDS may also be used as a programming aid by storing common source code, data to control software interfaces, data to control general maintenance and inquiry utilities and statistics to be used as a basis for the creation and maintenance of test files. The DDS is also the source for DML generation as well as the schema and subschema generation for DBMS.

E. ADVANTAGES OF USING DATA DICTIONARY SYSTEMS

A DDS benefits many users in an organization. Allen, et. al. [Ref. 28] identifies the following user groups and the DDS

functions of their prime interest.

(1) DBAs
who use the system as a major tool for inventorying the data resource, implementing standards, and designing monitoring and restructuring databases.

(2) Application Personnel
(e.g., analysts and programmers) who use the system to reduce program coding efforts, store the designs of evolving systems and support analysis of system changes.

(3) Operations Staff
who retrieve information about jobs from the database.

(4) DP Management
who receive high-level impact and summary reports about data usage from the database.

(5) End Users
who obtain descriptions of their data views from the database.

(6) Auditors
who examine system documentation provided through the database.

A DDS impacts upon the management of an organization by improving management's control and knowledge of the data resource. This control and knowledge is achieved by centralizing all information about the data in a convenient tool--the DDS. Some of the advantages to using a DDS in an organization include the following aspects:

(1) enables management to enforce data definition standards

(2) eliminates unwanted data redundancy

(3) aids in securing sensitive data

(4) assists management in quickly determining impacts of proposed changes to a system

(5) assists management in ensuring complete and accurate changeovers in the implementation of new systems

(6) supplies information about the creation, usage and relationships of data

(7) reduces the clerical load of a DBA

(8) gives a DBA control over design and use of a database by:
  (a) controlling and documenting formulation, meaning and usage of data structures
  (b) evaluating and controlling data redundancy
  (c) providing accurage data definitions for programs

(9) aids in the analysis of an organization's data flow by providing a method to track documents which flow through an organization

(10) provides a central source of information for designers to prevent redundancy and inconsistency in system design

(11) generates test data for designers

(12) provides documentation on systems design

(13) enforces data definition standards during program coding

(14) automatically generates code

(15) improves accuracy of finished programs by generation of test data and checking results automatically

(16) provides cross-referencing to assist in implementing approved changes to a system

(17) automatically implements amendments to operational systems

(18) provides documentation on changes to a system

(19) aids operations personnel in the creation of job control language parameters

(20) determines the source of data (including invalid data)

The DDS will allow automation of documentation, program coding, test data creation and checking and auditing the system output. DDS, therefore, allows management and control of a critical organizational resource--data. As this is the goal of IRM, it follows that use of a DDS will substantially assist an organization in achieving effective and efficient IRM.

F. SELECTION/EVALUATION CRITERIA

When management decides to implement a DDS, the first consideration is whether to purchase commercially available software or develop one in-house. Due to the volatile nature of the field, it is highly probable that enhancements will improve and increase the effectiveness of the DDS. Additionally, designing a DDS for a specific implementation will require a great deal of effort, costing a large amount of money, especially if the system must be continually updated. Finally, the DDS used by an organization must be highly reliable. It is possible that undetected errors might be resident in an in-house developed DDS for longer periods of time than in a commercially available system. For these reasons, Lefkovits [Ref. 29] suggests that any DDS utilized by an organization should be a commercially available system.

Before initiating the selection process, an organization must determine if a DDS is justified based upon an economic analysis of costs and benefits or savings of implementing the system. As in any economic analysis, determining an actual

dollar value for savings or benefits may be extremely difficult and is subjective judgement. Fig. 12 lists some aspects of costs and savings/benefits which should be considered in the economic analysis. Fig. 12 is not an all inclusive list; management should determine actual cost/benefit categories to be considered applicable to their organization.

COSTS	BENEFITS/SAVINGS
Acquisition	System Design and Development
Data Administration Staff	System Maintenance
Hardware Costs	Data Redundancy
Start-up Costs	Database Creation
Data Collection Costs	Auditing Information Resources
Maintenance	Improved Communications
Application System Changes	
User Education	

Figure 12--Costs and Savings/Benefits of DDS

Selection of a DDS should be based upon who will use the system and how it will be used, rather than what is the most technologically innovative system available. Lefkovits [Ref. 30] suggests the following selection and evaluation process:

(1) Determine requirements; classify which requirements are mandatory and which are desirable features with an associated point scale indicating importance.

(2) Develop a list of features to be used in the evaluation of DDS.

(3) Determine a mapping from requirements to features; multiple mappings may be possible.

(4) Compare features provided by commercially available systems to each mapping to determine if a system qualifies for further consideration (i.e., possesses all mandatory features).

(5) Compare those systems which qualify for the degree of compliance of any available desirable features, assigning a point value.

(6) Sum point values assigned to desirable features of qualified systems to select the DDS which bests meets the requirements.

This process is not without risk, especially since subjective judgement on the part of management is involved. The wrong system may still be selected for many reasons, including determination of improper requirements, usually due to a lack of user involvement in the selection process; unnecessary features given high point values while mandatory features were given low point values, due to technical bias of selection team; inconsistent evaluation of the system, due to different members of the selection team evaluating different systems as well as a lack of a well-defined measurement method; and undue emphasis on features needed in the future, but not at the time of implementation, which could result in user dissatisfaction with an unnecessarily complex system.

Once a system is selected and implemented, it should be evaluated periodically to determine whether or not it is performing acceptably. Often the requirements of the organization will change, requiring a reevaluation of the DDS to determine if it meets the new and/or changed requirements of the

organization. If the DDS no longer meets these requirements in an acceptable fashion, a new system must be selected.

G. FUTURE DIRECTIONS

No commercially available DDS is presently capable of providing all functions envisioned for its use. This is not to say that the DDS currently available are worthless, just that there is room for a great deal of improvement. Curtice [Ref. 31] notes that it appears that the use of DDS is proliferating without benefit of appropriate standards, adequate discipline or fundamental principles and methodology. Some of the contributing factors include:

(1) lack of generally accepted standards, or even guidelines, for what constitutes a good data definition

(2) lack of clarity about which important characteristics of data should be recorded in a DDS

(3) lack of a recognized and useful definition of "data element"

(4) lack of accepted discipline of conceptual or logical database design

(5) controversy about the best model for the conceptual level description of a database

Areas where DDS are weak and require more development in future are a greater integration of DDS into actual software lift cycle management, more powerful query and analysis capabilities and redesign of user interfaces to make them more "user-friendly". However, the major topics for consideration in the future development of DDS are their use in distributed

networks, in mini- and microcomputer applications, and, above all, integration into IRM and evolution into an IRDS.

IV. DATA DICTIONARY SYSTEMS AND THE FEDERAL GOVERNMENT

A. CONGRESS

1. Introduction

Congress has enacted two key pieces of legislation which impact the implementation of DDS by agencies of the Federal government. The Brooks Act has an indirect effect, being mainly concerned with the acquisition of automatic data processing equipment (ADPE). The Paperwork Reduction Act of 1980, on the other hand, establishes IRM as a mandatory government management concept.

2. The Brooks Act

Public Law 89-306; 40 United States Code Section 759 Section III to the Federal Property and Administrative Services Act of 1949 is commonly known as "the Brooks Act" due to the sponsorship of Representative Jack Brooks (D-Tex). It was enacted 30 October 1965 and authorized and directed the Administrator of the General Services Administration (GSA) to

> coordinate and provide for the economic and efficient purchase, lease and maintenance of automatic data processing equipment by Federal agencies

Prior to this time, Federal agencies pursued a course of purchasing or leasing ADPE based upon individual needs, resulting in large amounts of money being spent. Congress noted the increased government spending on ADPE and moved to control the proliferation of ADP systems within the Federal government.

The Brooks Act was the first attempt on the part of Congress to exert some type of control over Federal ADP spending.

The Brooks Act tasks GSA with being the sole procurement agent for the Federal government for all ADP acquisitions, which authority may be delegated in situations deemed necessary to affect efficient implementation. GSA was also tasked with managing a pool of equipment which could be transferred among various Federal agencies. The National Bureau of Standards (NES) was tasked with developing uniform Federal ADP standards to attempt to standardize Federal ADP operations. Finally, the Office of Management and Budget (OMB) was designated as policy maker and "referee" between GSA and user agencies in those cases of disagreement over the necessity of ADPE procurement.

The Brooks Act, which was enacted prior to the emergence of software as a major portion of the cost of a computer system (see Fig. 3), specifically states its applicability to ADP hardware and hardware maintenance services. However, the increasing availability and cost of software and software maintenance services are making a notable impact upon acquisition of new or upgraded computer systems, causing some reevaluation of the 1965 position. Commercially available software, which includes DDS, are now considered to be included in the provisions of the Brooks Act, and must, therefore, be purchased, leased or maintained efficiently and economically.

3. <u>The Paperwork Reduction Act of 1980</u>

Public Law 96-511; 44 United States Code Section 35, known as the Paperwork Reduction Act of 1980, was enacted 11 December 1980 in order to reduce paperwork and enhance the economy and efficiency of the Government and the private sector by improving Federal information policymaking. Representative Jack Brooks, best known for his sponsorship of the Brooks Act, was also instrumental in the enactment of the Paperwork Reduction Act of 1980. The stated purpose of the act is:

(1) to minimize the Federal paperwork burden for individuals, small businesses, State and local governments, and other persons;

(2) to minimize the cost to the Federal Government of collecting, maintaining, using and disseminating information;

(3) to maximize the usefulness of information collected by the Federal Government;

(4) to coordinate, integrate and, to the extent practicable and appropriate, make uniform Federal information policies and practices;

(5) to ensure that automatic data processing and telecommunications technologies are acquired and used by the Federal Government in a manner which improves service delivery and program management, increases productivity, reduces waste and fraud, and, wherever practicable and appropriate, reduces the information processing burden for the Federal Government and for persons who provide information to the Federal Government; and

(6) to ensure that the collection, maintenance, use and dissemination of information by the Federal Government is consistent with applicable laws, relating to confidentiality, including section 552a of title 5, United States Code, known as the Privacy Act.

The act further specifically defines data element and data element dictionary. A data element means a distinct piece of information such as a name, term, number, abbreviation or symbol, while a data element dictionary means a system containing standard and uniform definitions and cross references for commonly used data elements.

The Office of Information and Regulatory Affairs (OIRA) is a new office established by the Act within OMB. The Director of OIRA is responsible for developing and implementing Federal information policies, principles, standards and guidelines, acting as a focal point for Federal information management policy. Included in the general information policy functions of the Director is the development and implementation of uniform and consistent IRM policies and the evaluation of Federal agency information management practices to determine their adequacy and efficiency and to determine compliance of these practices with the policies, principles, standards and guidelines promulgated by the Director.

A stated goal of the Director of OIRA under the Act is to reduce the existing burden of Federal collection of information by 15% by 1 October 1982, and to reduce the existing burden by an additional 10% by 1 October 1983. Additionally, the Director was tasked to establish the Federal Information Locator System, establish standards and requirements of agency audits of all major information systems, identify areas of

duplication in information collecting and develop a schedule and methods for reducing this duplication by 1 October 1981. Finally, the Director was to develop, in consultation with the Administrator of GSA, a five-year plan for meeting the ADP and telecommunications needs of the Federal Government by 1 October 1982.

Under the Paperwork Reduction Act of 1980 each Federal agency is responsible for carrying out information management activities in an efficient, effective and economical manner and for complying with the information policies, principles, standards and guidelines prescribed by the Director of OIRA. Each Federal agency is required to designate a senior official or officials who report directly to the agency head to carry out the IRM responsibilities of the agency required by the Act. The Director of OIRA is tasked with the selective evaluation at least once every three years of the information management activities of each Federal agency to ascertain their adequacy and efficiency.

A key part of the Act is the establishment of the Federal Information Locator System in the OIRA. The Act envisions this system to be composed of a directory of information resources, a data element dictionary and an information referral service. The system is to serve as the authoritative register of all information collection requests (i.e., documents calling for the collection of information) or a centralized

listing of data available to Federal agencies. The system will promote data sharing and reduce data redundancy within the Federal government. OMB is presently testing a system based upon the Information Requirements Control Automation System (IRCAS) of the Office of the Secretary of Defense (OSD). IRCAS was designed to give OSD control over reports in an effort to eliminate duplication of information gathering. The system has been refined and updated and is presently being tested for possible implementation as the basis for the Federal Information Locator System. Testing will continue until at least March 1985.

The Paperwork Reduction Act of 1980 is a direct result of the recognition of Congress of the IRM concept and the necessity of implementing it to benefit the Federal Government. The key to IRM is to have the tools to manage information cheaply and effectively. The major tool to effect this management is DDS.

B. NATIONAL BUREAU OF STANDARDS

To utilize computer technology most effectively, it is desirable, to the extent feasible, to establish standards that are designed to achieve the maximum degree of compatibility and interchangeability among information systems. Federal agencies are required to implement and comply with the standards unless otherwise justified. This approach has far reaching and lasting benefits. From a management standpoint, the interchangeability

of equipment, programs and data throughout the entire Federal establishment would extend the efficiency and usefulness of Federal information systems, facilitate this orderly replacement as required and reduce the overall cost.

One of the provisions of the Brooks Act was the tasking of the Secretary of Commerce with providing Federal agencies with scientific and technological advisory services relating to ADP and related systems, and to make appropriate recommendations to the President relating to the establishment of uniform Federal ADP standards. The Brooks Act further authorized the Secretary to undertake any necessary research in the sciences and technologies of ADP computer and related systems required to support the duties assigned to the Secretary. OMB promulgated policy guidance to the Secretary of Commerce for the implementation of the Brooks Act. This guidance identified five areas for specific actions:

(1) Advisory and consulting services

(2) Development of voluntary commercial standards

(3) Recommendation for uniform Federal standards

(4) Research on Computer Science and Techniques

(5) Computer Services

The Secretary exercises his technical and scientific advisory role through the NBS. NBS provides this support through the programs of the NBS Institute for Computer Sciences and Technology (ICST), which was established in 1966 in response to the new responsibilities assigned to NBS under the Brooks Act.

ICST's long range plan calls for the development of standards and guidelines needed by Federal agencies to address the major problems of ADP use: to reduce the high costs

to reduce the high costs of software development and maintenance and to improve software quality;

to encourage the more efficient use and interchange of data

to better ADP operations, especially the security and integrity of operations; and

to improve capabilities for interconnecting components, systems and networks.

These standards are promulgated, through GSA, as Federal Information Processing Standards (FIPS) and collectively constitute the FIPS Register. All Federal agencies should establish and maintain a FIPS PUB/FIPS Register in accordance with FIPS PUB 0, "General Description of the Federal Information Processing Standards Register, 1 November 1968." Appendix A contains a listing of FIPS which have been published as of 31 March 1983 (FIPSPUB99). Overall, FIPS aid Federal agencies in three problem areas of computer compatibility (standard coding and data transfer), management and documentation, and security. FIPS are categorized as follows:

General Standards

Hardware Standards

Character Recognition
Interchange Codes and Media
Transmission
Interface
Data Entry Equipment
Computer Output Microfilm Readers

Software Standards

Programming Languages
Operating Systems
Operating Procedures
Media and Data Files
Data Management Applications
Software Engineering

Federal General Data Standards

Data Elements
Representations and Codes
Data Formats

ADP Operations Standards
Computer Security
Benchmarking for Computer Selection
Computer Performance Mangement
Management of Multivendor ADP Systems

In previous years ICST's technical assistance and research activities were limited to the direct support of standards development. However, recently ICST is beginning to research areas of increasing importance in Federal computer applications. Two major areas of interest are database technology and local area communications networking. In the area of database technology, ICST researchers are developing ways to express and manipulate the complex data structures involved in DBMS, DDS and other information processing systems which are used by Federal agencies to manage and control their data resources and to provide the capability of data sharing among many users.

The Federal Information Processing Standards Coordinating and Advisory Committee (FIPSCAC) coordinates the work assignments of a series of FIPS Task Groups which are established to

study specific topics relative to establishment of standards. Appendix B lists the various FIPS Task Groups. The draft proposals developed by FIPS Task Groups are reviewed by the FIPSCAC. The FIPSCAC also serves as a general advisory group to the Department of Commerce on information processing standards and advises on current and emerging issues relating to ADP standards. Each FIPS Task Group is composed of technical personnel with a knowledge of their particular Federal agency's requirements. These personnel assist NBS in matters relating to the development, adoption and implementation of standards and in providing better coordination of the Federal ADP Standards Program.

In May 1974, the Comptroller General of the United States, in a report to the Congress, noted that the cost for Federal data collection and data handling activities was estimated to exceed $5 billion annually [Red. 32]. There is, therefore, a great deal of pressure to reduce redundant data resources, and improve the utility of existing data resources. DDS is an appropriate tool for use by Federal agencies to eliminate unnecessary data gathering, reduce costs, and improve information systems' effectiveness. NBS established FIPS Task Group 17 in order to develop guidelines for constructing DDS and to identify relevant performance characteristics of the automated processes designed to use and maintain DDS. This Task Group produces two reports, NBS Special Publication 500-3, "Technical Profile of Seven Data Element Dictionary/Directory Systems," and NBS

Special Publication 500-16, "A Survey of Eleven Government-Developed Data Element Dictionary/Directory Systems," in 1977. Further research in this area by this Task Group resulted in the publication of FIPS PUB 76, "Guideline for Planning and Using a Data Dictionary System" of 20 August 1980. This guideline provides assistance to Federal ADP Management and technical staff in planning and using DDS, describing the capabilities of a DDS, discusses selection considerations, and provides guidance for preimplementation planning, implementation, and operational use of a DDS. It is to serve as the basic reference document for general use by Federal agencies in the implementation and use of a DDS.

ICST is also engaged in a series of Database Directions Workshops in conjunction with the Association for Computing Machinery (ACM). The first workshop, held October 1975, was concerned with database fundamentals--language structures, standards needed to govern future growth and benefits to be expected from the database environment. The second workshop, held 1-3 November 1977, addressed the conversion problem inherent in adjusting from one database environment to another. The third workshop, held 20-22 October 1980, focused upon strategies and tools for implementation of IRM. This workshop dealt primarily with DDS and their effective use as the major tool to implement IRM. Based upon the discussions of this third workshop, it would appear that the requirements of IRM go beyond

the capabilities of the currently available DDS. The evolution of DDS into IRDS for support of IRM, which is the current trend in the marketplace, was recognized by the workshop. It would appear that the next step would be research of the IRDS for possible publication as a FIPS.

C. DEPARTMENT OF THE NAVY

The Department of the Navy (DON), as an agency of the Federal Government, is bound by legislative and executive policy. Therefore, whenever Congress enacts legislation affecting Federal agencies, the Executive offices promulgate policy for those Federal agencies affected.

The Brooks Act, having been in effect for over nineteen years, has given rise to a plethora of regulations governing Federal ADP management and procurement. Executive regulations which have been promulgated in response to the Brooks Act include the Federal Property Management Regulations, the Federal Procurement Regulations and Federal Management Circular 74-5 issued by GSA; eight OMB Circulars (including Circular A-71 and A-75); various reports and studies published by the General Administration Office (GAO); and the FIPS published by NBS.

GSA is the major agency affecting ADP acquisition procedures, with additional guidance provided by OMB and NBS. This guidance provides the framework within which the Department of Defense (DOD) and DON must operate. Within DOD there are a multitude

of regulations governing ADP management and procurement, chief among them DOD Directive 4105.00, "Selection and Acquisition of Automatic Data Processing Resources," and DOD Instruction 5100.40, "Responsibility for the Administration of the DOD Automatic Data Processing Program." DON, in turn, has implemented regulations promulgating this policy within the Navy. The Secretary of the Navy (SECNAV) has over forty instructions in effect, the most important of which is SECNAV Instruction 5236.1A, "Specification, Selection, and Acquisition of Automatic Data Processing Equipment." At the next lower level in the hierarchy, the Chief of Naval Operations (CNO) or OPNAV level, there are over 35 instructions containing information regarding ADP management specifically applied to the Navy.

As can be seen by the vast numbers of regulations implemented at each level of the hierarchy from Congress to DON, ADP acquisition and management is viewed very seriously by the Federal Government. This desire for effective and efficient management and control of ADP resources and the recognition of the newly emerging concept of IRM by the Federal Government has led to further legislation in the form of the Paperwork Reduction Act of 1980. As with any Congressional enactment covering a broad topic involving many hierachical levels in the Federal Government, there is some lag between Congressional action and Federal agency implementation.

CNO directed the establishment of the Information Management Division (OP-945) as of 1 August 1983 [Ref. 33]. He also effected an organizational realignment of functions and resources of the Navy Records and Information Management Division (OP-09B1), which was disestablished 15 January 1984 [Ref. 34]. The purpose of the realignment was to provide for increased attention by the Navy to information systems management, including a shift of emphasis from ADP hardware and software to IRM. Under the direction of OP-094, Command and Control, OP-945 is responsible for the development of program policy. Commander, Naval Data Automation Command (COMNAVDAC) is responsible for program execution, as assigned by CNO, upon development of a strategic implementation plan by OP-945. COMNAVDAC is to submit an update to OPNAVINST 5450,200, "Mission and Functions of COMNAVDAC," reflecting the establishment of OP-945 for CNO review by 1 March 1984. The contents of OPNAVNOTE 5430 of 11 January 1984 will be incorporated into the OPNAV Organization Manual in the near future.

CNO also directed the revisions of OP-945 mission and functions. The revised mission is:

> To ensure optimum Navy information systems--ashore and afloat, combat and support--by providing policy, guidance, planning, standards, and assessment and to serve as Director, Department of the Navy Information Resources Management in direct support of the senior official designated in accordance with the Paperwork Reduction Act of 1980 (PL 96-511) [Ref. 35].

In support of this mission, 24 functions are identified for performance by OP-945 (Appendix C).

The actual strategic implementation of the mission and functions of OP-945 is in the process of being drafted. Upon CNO approval of OP-945 strategic plan to establish a Navy-wide IRM policy, COMNAVDAC will be responsible for execution. It appears most probable that a DDS will be part of the strategic plan, in direct support of the function to register and standardize data elements. DDS could also support the effective and efficient use of information systems technology in support of DON missions, validation of information requirements, and development of information methods and techniques.

It can be seen that with the passage of the Paperwork Reduction Act of 1980 and the establishment of OP-945 that the Federal Government and the Navy have fully accepted and endorsed IRM. Implementation of IRM is vital to ensuring successful and integrated DP operations. As a key to the success of IRM within an organization, DDS will also play a vital role in the future of DP within the Navy.

V. CONCLUSIONS

The advances made in DP since the introduction of the first general purpose computers in 1951 have led to an explosive proliferation of computer usage in the last ten years. As more and more operations vital to an organization become automated, the actual processing of data and the information which is produced from it become critical to the operation of that organization. In order to effectively and efficiently manage and control information, as it would any other critical organizational resource, management must implement and support IRM. One tool which management can utilize to effect IRM is the DDS.

DDS are presently in an evolutionary state. DDS implementation has lagged somewhat behind that of the earlier developed DBMS, and the confusion regarding scope, definition and integration of currently available DDS somewhat hinders the effective and widespread implementation of DDS. Many functions which DDS purport to possess are largely theoretical in nature. System complexities and lack of user education often lead to erroneous or misdirected use of DDS, in those instances where they are present. However, increasing interest and attention in this area has led to improvements in DDS as well as their potential for development into even more complex IRDS. The IRDS, presently at the theoretical stage, would effect even greater support of IRM within an organization.

IRM is a relatively recent innovation which is presently revolutionizing the DP environment. Recognition of IRM as an essential component of the successful management of an organization has even reached agencies of the Federal Government. Without IRM, no organization will be able to effectively function in the future DP environment. Without DDS, no organization will be able to effectively implement IRM.

APPENDIX A

LISTING OF FEDERAL INFORMATION PROCESSING STANDARDS (FIPS)

I. GENERAL

General Description of the Federal Information Processing Standards Register
FIPSPUB0   1 November 1968

Federal Information Processing Standards Index
FIPSPUB12-2   1 December 1974

Objectives and Requirements of the Federal Information Processing Standards Program
FIPSPUB23   15 February 1973

Standardization of Data Elements and Representations
FIPSPUB28   5 December 1973

Interpretation Procedures for Federal Standard COBOL
FIPSPUB29   30 June 1974

Guide for the Use of International System of Units (SI) in Federal Information Processing Standards Publications
FIPSPUB34   1 January 1975

Guide for the Implementation of Federal Information Processing Standards (FIPS) in the Acquisition and Design of Computer Products and Services
FIPSPUB80   19 December 1980

II. HARDWARE STANDARDS

A. Character Recognition

Optimal Character Recognition Character Sets
FIPSPUB32-1 25 June 1982

Character Set for Handprinting
FIPSPUB33 1 October 1974

Guideline for Optical Character Recognition Forms
FIPSPUB40 1 May 1976

Optical Character Recognition (OCR) Inks
FIPSPUB85 7 November 1980

Optical Character Recognition (OCR) Character Positioning
FIPSPUB89 4 September 1981

B. Interchange Codes and Media

Code for Information Interchange
FIPSPUB1-1 24 December 1980

Perforated Tape Code for Information Interchange
FIPSPUB2 1 November 1968

Recorded Magnetic Tape for Information Interchange
(800 CPI, NRZI)
FIPSPUB3-1 1 November 1968

Implementation of the Code for Information Interchange and
Related Media Standards
FIPSPUB7 7 March 1969

Rectangular Holes in 12-Row Punched Cards
FIPSPUB13 1 October 1971

Hollerith Punched Card Code
FIPSPUB14-1 24 December 1980

Subsets of the Standard Code for Information Interchange
FIPSPUB15 1 October 1971

Recorded Magnetic Tape for Information Interchange (1600
CPI, Phase Encoded)
FIPSPUB25 30 June 1973

One-Inch Perforated Paper Tape for Information Interchange
FIPSPUB26   30 June 1973

Take-Up Reels for One-Inch Perforated Tape for Information Interchange
FIPSPUB27   30 June 1973

Code Extension Techniques in 7 or 8 Bits
FIPSPUB35   1 June 1975

Graphic Representation of the Control Characters of ASCII (FIPSPUB1)
FIPSPUB36   1 June 1975

Recorded Magnetic Tape for Information Interchange, 6250 CPI (246 CPMM), Group Coded Recording
FIPSPUB50   1 February 1978

Magnetic Tape Cassettes for Information Interchange (3.810 MM [0.150 IN] Tape at 32 BPMM [800 BPI], Phase Encoded)
FIPSPUB51   1 February 1978

Recorded Magnetic Tape Cartridge for Information Interchange 4-Track, 6.30 MM (¼ IN), 63 BPMM (1600 BPI), Phase Encoded
FIPSPUB52   15 July 1978

Computer Output Microform (COM) Formats and Reduction Rations, 16 MM and 105 MM
FIPSPUB54   15 July 1978

Guideline for Inspection and Quality Control for Alphanumeric Computer-Output Microforms
FIPSPUB82   26 September 1980

Magnetic Tape Cassettes for Information Interchange, Dual Track Complementary Return-to-Bias (CRB) Four-States Recording on 3.81 MM (0.150 IN) Tape
FIPSPUB91   12 March 1982

Parallel Recorded Magnetic Tape Cartridge for Information Interchange, 4-Track, 6.30 MM (¼ IN), 63 BPMM (1600 BPI), Phase Encoded
FIPSPUB93   29 June 1982

C. Transmission

Bit Sequencing of the Code for Information Interchange in Serial-by-Bit Data Transmission
FIPSPUB16-1 1 September 1977

Character Structure and Character Parity Sense for Serial-by-Bit Data Communication in the Doce for Information Interchange
FIPSPUB17-1 1 September 1977

Character Structure and Character Parity Sense for Parallel-by-Bit Data Communication in the Code for Information Interchange
FIPSPUB18-1 1 September 1977

Synchronous Signaling Rates Between Data Terminal and Data Communication Equipment
FIPSPUB22-1 1 September 1977

Synchronous High Speed Data Signaling Rates Between Data Terminal Equipment and Data Communications Equipment
FIPSPUB37 15 June 1975

Advanced Data Communication Control Procedures (ADCCP)
FIPSPUB71 14 May 1980

Guideline for Implementing Advanced Data Communication Control Procedures (ADCCP)
FIPSPUB78 26 September 1980

D. Interface

I/O Channel Interface
FIPSPUB60-1 27 August 1979

Channel Level Power Control Interface
FIPSPUB61-1 13 July 1982

Operational Specifications for Magnetic Tape Subsystems
FIPSPUB62 16 February 1979

Operational Specifications for Rotating Mass Storage Subsystems
FIPSPUB63-1 14 April 1983

Operational Specifications for Fixed Block Rotating Mass Storage Subsystems
FIPSPUB97 4 February 1983

E. Data Entry Equipment

Guideline for Selection of Data Entry Equipment
FIPSPUB67 30 September 1979

F. Computer Output Microfilm Readers

Microfilm Readers
FIPSPUB84 31 October 1980

III. SOFTWARE STANDARDS

A. Programming Language

COBOL
FIPSPUB21-1   1 December 1975

Interpretation Procedures for Federal Standard COBOL
FIPSPUB29   30 June 1974

Aid for COBOL Program Conversion (FIPSPUB21 to FIPSPUB21-1)
FIPSPUB43   1 December 1975

Federal Standard COBOL Pocket Guide
FIPSPUB47   1 February 1977

Minimal BASIC
FIPSPUB68   4 September 1980

FORTRAN
FIPSPUB69   4 September 1980

B. Operating Systems

C. Operating Procedures

Magnetic Tape Labels and File Structure for Information Interchange
FIPSPUB79   17 October 1980

D. Documentaion

Dictionary for Information Processing
FIPSPUB11-1   30 September 1977

Guidelines for Describing Information Interchange Formats
FIPSPUB20   1 March 1972

Flowchart Symbols and Their Usage in Information Processing
FIPSPUB24   30 June 1973

Software Summary for Describing Computer Programs and Data Systems
FIPSPUB30   30 June 1974

Guidelines for Documentation of Computer Programs and Automated Data Systems
FIPSPUB38 15 February 1976

COBOL Coding Form
FIPSPUB44 1 September 1976

Transmittal Form for Describing Computer Magnetic Tape File Properties
FIPSPUB53 . 1 April 1978

Guidelines for Documentation of Computer Programs and Automated Data Systems for the Initiation Phase
FIPSPUB64 1 August 1979

E. Media and Data Files

Code for Information Interchange
FIPSPUB1-1 24 December 1980

Message Format for Computer-Based Message Systems
FIPSPUB98 1 March 1983

F. Data Management Applications

Guideline for Planning and Using a Data Dictionary System
FIPSPUB76 20 August 1980

Guideline for Planning and Management of Database Applications
FIPSPUB77 1 September 1980

Guideline on Integrity Assurance and Control in Database Administration
FIPSPUB88 14 August 1981

G. Software Engineering

Guideline: A Framework for the Evaluation and Comparison of Software Development Tools
FIPSPUB99 31 March 1983

IV. FEDERAL GENERAL DATA STANDARDS

A. Data Elements

B. Representations and Codes

Calendar Date
FIPSPUB4 1 November 1968

States and Outlying Areas of the United States
FIPSPUB5-1 15 June 1970

Countries and County Equivalents of the States of the United States
FIPSPUB6-3 15 December 1979

Standard Metropolitan Statistical Areas (SMSA)
FIPSPUB8-4 30 June 1974

Congressional Districts of the United States
FIPSPUB9 14 November 1969

Countries, Dependencies, and Areas of Special Sovereignty
FIPSPUB10-2 15 November 1976

Guidelines for Registering Data Codes
FIPSPUB19 1 February 1972

Guide for the Development, Implementation, and Maintenance of Standards for the Representation of Computer Processed Data Elements
FIPSPUB45 30 September 1976

Guideline for Codes for Named Populated Places and Related Entities of the States of the United States
FIPSPUB55 5 February 1982

Representations of Local Time of the Day for Information Interchange
FIPSPUB58 1 February 1979

Representations of Universal Time, Local Time Differentials, and United States Time Zone References for Information Interchange
FIPSPUB59 1 February 1979

Standard Industrial Classification (SIC) Codes
FIPSPUB66   15 August 1979

Representation of Geographic Point Locations for Information Interchange
FIPSPUB70   24 October 1980

Guideline for Standard Occupational Classification (SOC) Codes
FIPSPUB92   24 February 1983

Codes for the Identification of Federal and Federally-Assisted Organizations
FIPSPUB95   23 December 1982

C. Data Formats

V. ADP OPERATIONS STANDARDS

A. Computer Security

Guidelines for Automatic Data Processing Physical Security and Risk Management
FIPSPUB31 June 1974

Glossary for Computer Systems Security
FIPSPUB39 15 February 1976

Computer Security Guidelines for Implementing the Privacy Act of 1974
FIPSPUB41 30 May 1975

Data Encryption Standard
FIPSPUB46 15 January 1977

Guidelines on Evaluation of Techniques for Automated Personal Identification
FIPSPUB48 1 April 1977

Guidelines for Automatic Data Processing Risk Analysis
FIPSPUB65 1 August 1979

Guidelines for Security of Computer Applications
FIPSPUB73 30 June 1980

Guideline on User Authentication Techniques for Computer Network Access Control
FIPSPUB83 29 September 1980

B. Benchmarking for Computer Selection

Guidelines for Benchmarking ADP Systems in the Competitive Procurement Environment
FIPSPUB42-1 15 May 1977

Guideline on Constructing Benchmarks for ADP System Acquisitions
FIPSPUB75 18 September 1980

C. Computer Performance Management

Guideline on Computer Performance Management: An Introduction
FIPSPUB49 1 May 1977

Guidelines for the Measurement of Interactive Computer Service Response Time and Turnaround Time
FIPSPUB57   1 August 1978

Guidelines for the Measurement of Remote Batch Computer Service
FIPSPUB72   1 May 1980

Guideline for Developing and Implementing a Charging System for Data Processing Services
FIPSPUB96   6 December 1982

D. Management of Multivendor ADP Systems

Guideline for Managing Multivendor Plug-Compatible ADP Systems
FIPSPUB 56   15 September 1978

APPENDIX B

FIPS TASK GROUPS

1 Objectives and Requirements for Standards

2 Data Terminals and Data Interchange System Requirements

3 Character Subsets, Sign Conventions and Packing Techniques

4 Subsections on Standards for Use in Requests for Proposals

5 Federal Information Processing Vocabulary

6 Computer Magnetic Tapes

7 Magnetic Tape Labels for Information Interchange

8 Guidelines for Describing Data Interchange Formats

9 COBOL Standards

10 Guidelines for Computer System and Component Performance Evaluation

11 Optical Character Recognition

12 Significance and Impact of ASCII as a Federal Standard

13 Workload Definition/Benchmarking

14 Documentation for Information Processing Systems

15 Computer Systems Security

16 Basic Standard Programming Language

17 Data Element Dictionary

APPENDIX C

FUNCTIONS OF OP-945

1. Serves as principal advisor to OP-094 on all matters pertaining to information systems resources including: information resources management, information requirements, information and office systems, embedded computer resources, mission critical computers, data processing, records and forms management, postal affairs, and computer security.

2. Supports the senior official designated in accordance with the Paperwork Reduction Act of 1980 (PL96-511) and the DON Senior ADP Policy Official.

3. Acts to encourage effective and efficient use of information systems technology in support of DON missions.

4. Maintains awareness of external policy and regulations impacting Division programs, influences development and modification of those policies and regulations to the extent appropriate, and assures DON compliance.

5. Develps DON and Navy policy, procedures, objectives, manuals, handbooks, criteria, and other issuances as needed for implementation of the Division programs.

6. Maintains awareness of DOD, Federal, industry, and academic developments and actions of potential concern to the Division and promulgates such information as appropriate.

7. Coordinates action on GAO, Congressional, internal audit, inspector general, and other reviews, surveys, and audits in areas of concern to the Division.

8. Represents the DON externally on matters concerning Division programs not related to specific information systems.

9. Represents the DON externally on matters concerning specific information systems.

10. Validates information requirements, assuring that they are justified and non-duplicative, and that effective information systems support is provided.

11. Develops the top level information systems architecture for the DON and the strategic information systems plan in support of that architecture.

12. Provides leadership to teams charged to develop information systems architecture for designated systems.

13. Monitors the development and implementation of information systems architecture and plans.

14. Reviews plans and project approval requests for compliance with architecture, appropriate interface provisions and general soundness of approach.

15. Assures maximum practicable standardization of information systems.

16. Serves as DON Assessment Sponsor for information systems and otherwise review, prepares, and defends Program Objectives Memorandum and budget Submissions as appropriate.

17. Serves as program coordinator for designated programs such as THAIS, STAIRS, Fleet Work Processing Program, SNAP, and AN/UYK-43/44.

18. Acts as designator advisor for designators and ratings covered by Division programs and sponsors a civilian career management program for related series.

19. Sponsors development and promulgation of information systems technical standards.

20. Sponsors development of new information methods and technology, including information requirements description techniques, and acts to obtain effective use of new developments in DON information systems.

21. Assures appropriate training for users of information systems.

22. Provides for the registration and standardization of data elements.

23. Assesses progress and status of Division programs at least annually.

24. Advises OP-094 on CNO command related matters affecting NAVDAC and serves as NAVDAC program coordinator.

LIST OF REFERENCES

1. O'Brien, J. A., Computers in Business Management: An Introduction, 3d ed., p. 23, Richard D. Irwin, Inc., 1982.

2. ibid., pp. 23-25.

3. Shooman, M. L., Software Engineering: Design, Reliability and Management, p. 11, McGraw-Hill, 1983.

4. Nolan, R. L. and Wetherbe, J. C., "Toward a Comprehensive Frameword for MIS Research," MIS Quarterly, v. 4, n. 2, pp. 33, June 1980.

5. O'Brien, op. cit., p. 660.

6. Kroenke, D. M., Database Processing: Fundamentals, Design, Implementation, 2d ed., pp. 2-4, Science Research Assoc., Inc., 1983.

7. ibid., p. 206.

8. ibid., pp. 11-13.

9. Somerville, I., Software Engineering, p. 4, Addison-Wesley, 1982.

10. U. S. Department of Commerce, National Bureau of Standards Special Publication 500-92, "Data Base Directions: Information Resource Management--Strategy and Tools," Goldfine, A. H., ed., p. 50, 1982.

11. U. S. Department of Commerce, National Bureau of Standards Special Publication 500-28, "Database Administration: Concepts, Tools, Experiences and Problems," Leong-Hong, B. and Marron, B., pp. 4-5, 1978.

12. Kroenke, op. cit., p. 193.

13. ibid., p. 200.

14. National Bureau of Standards Special Publication 500-28, op. cit., pp. 10-14.

15. Cardenas, A. F., Data Base Management Systems, p. 462, 1979.

16. National Bureau of Standards Special Publication 500-28, op. cit., pp. 127-128.

17. "The British Computer Society Data Dictionary Systems Working Party Report," Data Base, v. 9, n. 2, p. 3, Fall 1977.

18. U. S. Department of Commerce, National Bureau of Standards Special Publication 500-3, "Technical Profile of Seven Data Element Dictionary/Directory Systems," Leong-Hong, B. and Marron, B., February 1977.

19. Allen, F. W., Loomis, M. E. S. and Mannino, M. V., "The Integrated Dictionary/Directory System," ACM Computing Surveys, v. 14, n. 2, p. 247, June 1982.

20. National Bureau of Standards Special Publication 500-92, op. cit., p. 18.

21. National Bureau of Standards Special Publication 500-3, op. cit., p. 6.

22. Plagman, B. K., "Data Dictionary/Directory System: A tool for Data Administration and Control," Auerbach Information Management Services--Data Base Management, No. 22-01-02, pp. 3-4, 1977.

23. Allen, et. al., op. cit., p. 247.

24. ibid., p. 248.

25. ibid., pp. 250-254.

26. "The British Computer Society Data Dictionary Systems Working Party Report," op. cit., p. 3.

27. ibid., p. 5.

28. Allen, et. al., op. cit., pp. 247-248.

29. Lefkovits, H. C., Sibley, E. H. and Lefkovits, S. L., Information Resource/Data Dictionary Systems, p. 1-38, QED Information Sciences, 1983.

30. ibid., p. 1-39 to 1-41.

31. Curtice, R. M., "Data Dictionaries: An Assessment of Current Practices and Problems," Proceedings of the Seventh International Conference on Very Large Data Bases, p. 565, Arthur D. Little, Inc., 1981.

32. U. S. Department of Commerce, Federal Information Processing Standards Publication 76, "Guideline for Planning and Using a Data Dictionary System," p. ii, 20 August 1980.

33. OPNAV Notice 5430 Ser 09B24/312958 of 1 August 1983.

34. OPNAV Notice 5430 Ser 09B24/44308705 of 11 January 1984.

35. ibid.

BIBLIOGRAPHY

Allen, F. W., Loomis, M. E. S. and Mannino, M. V., "The Integrated Dictionary/Directory System," ACM Computing Surveys, v. 14, n. 2, June 1982.

"The British Computer Society Data Dictionary Systems Working Party Report," Data Base, v. 9, n. 2, Fall 1977.

Coulson, C. J., "People Just Aren't Using Data Dictionaries," Computerworld, v. 16, n. 33, 16 August 1982.

Curtice, R. M., "Data Dictionaries: An Assessment of Current Practice and Problems," Proceedings of the Seventh International Conference on Very Large Data Bases, 1981.

Curtice, R. M. and Dieckmann, E. M., "A Survey of Data Dictionaries," Datamation, v. 27, n. 3, March 1981.

Elbra, R. A. and Hardy, N., "Cost Justification of Data Dictionaries," Computer Bulletin, ser. 2, n. 25, September 1980.

Engelbrecht, R., "From Data Dictionary to System Dictionary," First IEEE Computer Society International Conference on Medical Computer Science/Computational Medicine, 1982.

Gillenson, M. L., "The State of Practice of Data Administration--1981," Communications of the ACM, v. 25, n. 10, Oct. 1982.

Henderson, M. M., "Compilation of Bibliographic Data Element Dictionaries," National Bureau of Standards Special Publication 500-94, 1982.

Johnston, C. I. and Stone, A. S., "Monitoring CODASYL Database Management Systems," Computer Performance, v. 3, n. 1, March 1982.

Kennedy, R. M., "Data Dictionary Experience at West Midlands Gas," Database Journal, v. 11, n. 1, 1981.

Kreitzer, L. W., "Data Dictionaries--the Heart of IRM," Infosystems, v. 28, n. 2, February 1981.

Lefkovits, H. C., Sibley, E. H. and Lefkovits, S. L., Information Resource/Data Dictionary Systems, QED Information Sciences, 1983.

Mathieson, R. M., "The Control and Audit of Proprietary Software," Computer Audit and Control-State of the Art Report, Infotech, 1980.

Morehen, H. L., "Data Dictionaries, Evaluation and Implementation," Australian Computer Bulletin, v. 6, n. 10, Nov 1982.

Patel, H., "The Data Dictionary as a Tool for Systems Auditability and Control," Computer Audit and Control-State of the Art Report, Infotech, 1980.

Plagman, B. K., "Data Dictionary/Directory Systems: A Tool for Data Administration and Control," Auerbach Information Management Services--Data Base Management, No. 22-01-02, 1977.

"Proceedings of the Conference on Data Dictionary Systems," Computer Bulletin, ser. 2, n. 18, December 1978.

Schussel, G., "The Role of Data Dictionaries," Datamation, v. 23, n. 6, June 1977.

Sibley, E. H., Scallan, P. G. and Clemons, E. K., "The Software Configuration Management Database," AFIPS Conference Proceedings, v. 50, 1981.

Smith, A. A., "Installing a Data Dictionary System," Ninth Australian Computer Conference, vol. 2, Australian Computer Society, 1982.

Swager, J. R., "Architecture of a Distributed Database Information Resource," AFIPS Conference Proceedings, v. 50, 1981.

Ting, T. C., "ADD: An Automated Tool for Program Design and Documentation," National Bureau of Standards Special Publication 500-94, 1982.

U. S. Department of Commerce, Federal Information Processing Standard Publication 76, "Guideline for Planning and Using a Data Dictionary System, 20 August 1980.

U. S. Department of Commerce, National Bureau of Standards Special Publication 500-3, "Technical Profile of Seven Data Element Dictionary/Directory Systems," Leong-Hong, B. and Marron, B., February 1977.

U. S. Department of Commerce, National Bureau of Standards Special Publication 500-16, "Survey of Eleven Government Developed Data Element Dictionary/Directory Systems," 1977.

U. S. Department of Commerce, National Bureau of Standards Special Publication 500-28, "Database Administration: Concepts Tools, Experiences and Problems," Leon-Hong, B. and Marron, B., 1978.

U. S. Department of Commerce, National Bureau of Standards Special Publication 500-92, "Data Base Directions: Information Resource Management--Strategy and Tools," Goldfine, A. H., ed., 1982.

Van Duyn, J., Developing a Data Dictionary System, Prentice-Hall, 1982.

Vanacek, M. G., Solòmon, I. and Mannino, M. V., "The Data Dictionary: An Evaluation From the EDP Audit Perspective," MIS Quarterly, v. 7, n. 1, March 1983.

Windson, A. T., ed., Using the ICL Data Dictionary. Proceedings of the User Group, Shiva Publishing, 1980.

Zahran, F. S., "The Extensibility Feature of Data Dictionary Systems," Proceedings of the First British National Conference on Databases, 1981.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria, Virginia 22314	2
2.	Library, Code 0142 Naval Postgraduate School Monterey, California 93943	2
3.	Computer Technology Curricular Office Naval Postgraduate School Code 37 Monterey, California 93943	1
4.	Professor Daniel R. Dolk, (Code 54Dk) Department of Administrative Sciences Naval Postgraduate School Monterey, California 93943	1
5.	LCDR Bill Johnson NSHS Bethesda Detachment AHS Building 902 Fort Sam Houston, Texas 78234	1
6.	LT. D. L. Robertson Naval Military Personnel Command (N16G) Department of the Navy Washington, D. C. 20307	1

207577

Thesis
R5992 Robertson
c.1 Data dictionary systems and their role in information resources management.